PUTTING THE DISTRIBUTED PRACTICE EFFECT INTO CONTEXT

CHRISTINA WESTON

A DISSERTATION SUBMITTED TO
THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN PSYCHOLOGY
YORK UNIVERSITY
TORONTO, ONTARIO

September 2018

Abstract

Spaced repetition leads to superior final memory relative to massed repetition, a phenomenon known as the distributed practice effect. However, when items are repeated in variable study contexts across learning opportunities (relative to a consistent study context), the advantage of distributed practice over massed practice is typically reduced. In this dissertation, the effect of study context on the distributed practice effect was investigated from a neural perspective (Study 1) and from a developmental perspective (Study 2). In Study 1, event-related potentials (ERPs) were recorded as participants learned stimuli repeated after massed or distributed lags on either a consistent or variable background. After a fixed retention interval, stimuli were presented for a third time and participants' recognition memory was tested. Behavioural evidence of a Lag × Study Context interaction was mixed. The ERP data revealed a neural distinction between massed and distributed repetitions during the study phase in terms of the late positivity component (LPC); however, the LPC was not further defined by the study context manipulation. During the test phase, distributed, variably studied repetitions engendered the greatest neural familiarity response compared to all other repetition conditions. The ERP results provided converging evidence in support of a study-phase retrieval explanation of the distributed practice effect, which would not have been obvious using behavioural measures alone. In Study 2, younger and older participants learned stimuli repeated after varying lags on either a consistent or variable background. The background scenes were either shared among all to-be-learned items (Experiment 2A) or unique to each to-be-learned item (Experiment 2B). After the study phase, participants' free recall memory was tested. Based on a theory suggesting that older adults have difficulty binding items with their respective study contexts, it was hypothesized that manipulations of study context would have less of an impact on the distributed practice effect in aging. Although older adults did have greater difficulty identifying whether a repeated item's study context had changed throughout the study phase, they still exhibited similar final recall performance to younger adults during the test phase. Comparing data from the two experiments, the results also revealed that variations to study context might actually enhance the distributed practice effect in certain learning situations. This enhancement effect, which warrants further investigation, might depend on the type of material being learned and/or the variety of contextual information available during study.

Keywords: Distributed practice effect, study context, event-related potentials, cognitive aging.

Acknowledgements

Reflecting on my time as a PhD student, I can say that completing this dissertation has been a “desirably difficult” experience... certainly one that I won’t forget![1] I am grateful to so many people who have supported me throughout my academic pursuit. First, I would like to thank my supervisor, Melody, for her mentorship and friendship over the last several years. Above all, she has instilled in me an open-mindedness, curiosity, and persistence in reaching my goals that I know will be useful wherever life takes me next. Thank you also to my committee members, Gary and Dale, for their valuable comments throughout the planning and writing of my research (and to Gary for his inspirational teaching during my TA years). My sincerest gratitude to Reno, a former post-doc in our lab, who had many thoughtful discussions with me about my research, trained me in event-related potentials, and assisted with programming. Thank you to my volunteers who helped with data collection—Orly, Heather, and Negar—and Andrée-Ann at Glendon for sharing resources and teaming up on recruitment efforts. A shout-out to all of my wonderful friends from lab, past and present—Linda, Denise, Irina, Vanessa, Annalise, Katie, Alice, Galit, Tammy, Eliana, Rebecca, Margot—I am grateful for our regular chats and many laughs. Thank you to my family—Mom, Dad, and Jules. I would not have been able to come this far without their unconditional love and support. Finally, I would like to thank my husband, Robert, who has been at my side throughout this whole experience and whose positivity and encouragement have kept me grounded. Onto our next life adventure... parenthood!

---

[1] If this doesn’t make sense to you quite yet, hopefully it will by p. 11!

Table of Contents

List of Tables

List of Figures

Introduction

It is often said, “practice makes perfect”. Whether you are trying to fine-tune your golf swing or learn a new language, the act of practicing—or repeating—recently learned information is essential to committing it to memory. Practice is the foundation of many different kinds of learning. Despite its apparent simplicity, the benefit of repeated practice on memory is a behavioural phenomenon that has enamored experimental psychologists for well over a century. In 1885, memory researcher Hermann Ebbinghaus chronicled his personal attempts at memorizing sequences of consonant-vowel-consonant trigrams (e.g., BAF, XOF, MEQ) and retrieving them from memory after various time intervals ranging from minutes to months. In his reflections on the outcome of repeated practice, Ebbinghaus (trans. 1913) wrote:

> “The series are gradually forgotten, but—as is sufficiently well known—the series which have been learned twice fade away much more slowly than those which have been learned but once.... With any considerable number of repetitions a suitable distribution of them over a space of time is decidedly more advantageous than the massing of them at a single time.” (Chapter 8, Sections 31 and 34)

Ebbinghaus was the first researcher to formally compare and contrast different schedules of practice on subsequent memory. His work paved the way for hundreds of studies on what is now referred to as the *distributed practice effect* or *spacing effect*, one of the most robust and reliable findings in memory research. Numerous reviews and meta-analyses exist on the topic of distributed practice, some of which date back to its initial popularity in the 1970s, and others that were published just a few years ago (Crowder, 1976; Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006; Delaney, Verkoeijen, & Spirgel, 2010;

Dempster, 1996; Hintzman, 1974; Janiszewski, Noel, & Sawyer, 2003; Küpper-Tetzel, 2014; Maddox, 2016; Toppino & Gerbier, 2014). These articles are a testament to cognitive psychologists' persistent intrigue with distributed practice, its theories, and applications. With over 130 years of accumulated research, it turns out that explaining how and why practice makes perfect is far from simple.

Defining the Distributed Practice Effect

The distributed practice effect refers to the finding that items repeated in a distributed or temporally spaced manner during learning are more successfully retrieved on a final test of memory compared to items repeated in succession during learning, despite equal amounts of total study time. Figure 1 illustrates the design of a standard distributed practice paradigm where an item is first studied (*S1*) and then repeated again for restudy (*S2*) after a manipulated amount of time. The amount of time between S1 and S2 may be on the order of seconds or minutes (denoted *lag* and usually filled with other intervening to-be-learned items) or hours, days, months, and even years (denoted *interstudy interval).* In a typical experiment, a participant restudies a to-be-learned item after either a massed lag/interstudy interval (usually zero) or a spaced lag(s)/interstudy interval(s). After another period of time (denoted *retention interval*), the participant is tested for his/her memory of all items (*final test*). The final test may take the form of free recall, recognition, cued recall, or frequency judgement. The distributed practice effect is most commonly studied using *within-session* paradigms (e.g., comparing lags of 0, 4, 8 intervening items on an immediate test of memory). Although not the focus of the present dissertation, research has also explored distributed practice benefits using *between-session* paradigms (e.g., comparing interstudy intervals of 5 minutes, 1 day, and 1 week on a test two weeks later; e.g., Cepeda, Coburn, Rohrer, Wixted, Mozer, & Pashler, 2009). Furthermore, some studies have investigated the added benefit and scheduling of a third study opportunity (S3) in both within-session paradigms (e.g., Cull, Shaughnessy, & Zechmeister, 1996; Landauer & Bjork, 1978) and between-session paradigms (e.g., Küpper-Tetzel, Kapler, & Wiseheart, 2014).

The distributed practice effect has been observed across people of all ages, including: infants (e.g., Rovee-Collier, Evancio, & Earley, 1995), elementary school children (e.g., Sobel, Cepeda, & Kapler, 2011), high school students (e.g., Bloom & Schuell, 1981), college students (the most widely studied; e.g., Thios & D'Agostino, 1976), and middle-to-older aged adults (e.g., Balota, Duchek, & Paullin, 1989). It has even been observed in nonhuman species, for example, behavioural conditioning in honeybees (Menzel, Manz, Menzel, & Greggers, 2001). Distributed practice generalizes to a wide range of stimuli. For example, researchers have reported distributed practice benefits when testing rote memory for: words (e.g., Glenberg & Lehmann, 1980), word pairs (e.g., Madigan, 1969), trivia facts (e.g., Cepeda, Vul, Rohrer, Wixted, & Pashler, 2008), text passages (e.g., Rawson & Kintsch, 2005), objects (e.g., Paivio, 1974), and faces (e.g., Carpenter & DeLosh, 2005). The effect has been observed for motor skills activities such as: learning a finger tapping sequence (e.g., Shea, Lai, Black, & Park, 2000), playing a musical instrument (e.g., Simmons, 2012), practicing sports skills (e.g., Dail & Christina, 2004), playing a video game (e.g., Metalis, 1985), and medical students practicing surgical skills (e.g., Moulton et al., 2006). It has also been shown to improve complex reasoning skills such as category induction (e.g., Kornell & Bjork, 2008) and mathematical problem solving (e.g., Rohrer & Taylor, 2007). On the basis of its flexibility across people and domains, and its robustness (approximately $d = 1.0$ for verbal memory, according to Cepeda et al., 2006), an emerging literature is exploring applications of distributed practice in real-world education and training programs (Kapler, Weston, & Wiseheart, 2015; Wiseheart, Küpper-Tetzel, Weston, Kim, Kapler, & Foot, in press), and memory rehabilitation, such as for cases of amnesia (e.g., Green, Weston, Wiseheart, & Rosenbaum, 2014) and dementia (e.g., Cherry & Simmons-D'Gerolamo, 2005).

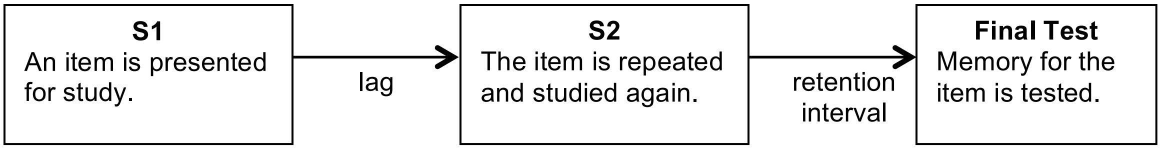

Figure 1. A standard distributed practice paradigm. An item is studied once (S1) and restudied (S2) after either a massed lag (zero intervening items) or a distributed lag(s) (one or more intervening items). After a fixed amount of time (the retention interval), memory for all items is tested. A distributed practice effect describes the finding that items repeated at a distributed lag(s) are remembered better on the final memory test than items repeated at a massed lag.

Explaining the Distributed Practice Effect

Three major classes of theories have been proposed to explain the distributed practice effect, including: deficient processing (e.g., Hintzman, 1974), encoding variability (e.g., Glenberg, 1979), and study-phase retrieval (e.g., Thios & D'Agostino, 1976). A review of the three theories demonstrates that they are not mutually exclusive nor can they each independently account for the enormous amount of published data across different stimuli, task demands, and timelines. Nowadays, researchers tend to subscribe to a hybrid account of the distributed practice effect—one that includes mechanisms from all three classic theories each contributing different weights depending on the paradigm (for a discussion, see Toppino & Gerbier, 2014). A hybrid account is an attractive explanation for the distributed practice effect because of its flexibility in accommodating the wide range of distributed practice benefits reported in the literature.

Deficient processing theory. Deficient processing theories (Hintzman, 1974) explain the distributed practice effect in terms of poor processing of massed repetitions compared to distributed repetitions during study. Inferior processing of massed repetitions may occur in a number of different ways. For example, some researchers have theorized that the second presentation (S2) of a massed item occurs before its first presentation (S1) has undergone full consolidation, thereby interrupting S1 processing of massed items (Landauer, 1969; Peterson, 1966). Other researchers have proposed that S2 of a massed item occurs so quickly after S1 that the learner does not have enough time for overt mental rehearsal of S1 (Rundus, 1971). Although these two theoretical propositions come to different conclusions about learner control (involuntary vs. voluntary, respectively), they similarly explain the distributed practice effect as inferior processing of massed items *at S1*.

Further exploration of deficient processing mechanisms in distributed practice indicates that the effect is more likely explained by inferior processing of massed items *at S2*. For example, in a study by Hintzman, Block, and Summers (1973; Exp 2), participants studied words once or twice (lags of 0, 1, 5, and 15 intervening items) in either the same modality or a different modality at S1 and S2 (visual-visual; auditory-auditory; visual-auditory; auditory-visual). In an immediate frequency judgment memory test, participants saw a list of targets and distractors and were asked to indicate whether each word had appeared once, twice, or never during the study phase. For words that they reported studying once or twice, they were also asked to indicate in which modality (or modalities) they remembered the word appearing. The results of this study revealed that participants more accurately identified the frequency of distributed words regardless of learning modality. Moreover, participants were more likely to (incorrectly) report that a massed item occurred only once and that it was (correctly) associated with the S1 learning modality. This finding provides evidence that participants did not completely reprocess massed items at S2, in turn, compromising performance for these items on the frequency judgment test. Contrary to deficient processing explanations at the time, processing of massed items at S1 was actually satisfactory. Similar results have been reported in subsequent research exploring whether deficient processing of massed items at S2 is involuntary or voluntary. Some evidence suggests that a learner may fail to reach a neurological response threshold for massed items at S2, analogous to a priming mechanism, thereby impoverishing S2 processing (Challis, 1993; Magliero, 1983; Russo, Parkin, Taylor, & Wilks, 1998), while other evidence indicates that a learner may choose to allocate less attention to a massed item at S2 because it was just processed at S1 (Shaughnessy, Zimmerman, & Underwood, 1972).

Encoding variability theory. Encoding variability theories offer another explanation of the distributed practice effect that focus on encoding processes during study. It is considered to be a descendent of the stimulus sampling theory of memory first presented by Estes (1955) and further discussed by others (e.g., Bower, 1972). Stimulus sampling theories state that a learning environment consists of abstract contextual elements that fluctuate with the passing of time, moving gradually and randomly from available to unavailable states of awareness. Contextual elements are defined as external characteristics of the environment (e.g., noise, lighting, presence of the experimenter) or internal characteristics of the learner (e.g., prior schemas, motivation, strategy). When an item undergoes study, it is encoded alongside a unique combination of contextual elements available at that time. Encoding variability is induced when an item undergoes restudy and is, again, encoded with a new combination of contextual elements. Traditional encoding variability theories treat each encoding opportunity as an independent memory trace. In the case of repetition, when an item is restudied at S2, some contextual elements will be new (different than S1) while others are expected to be old (same as S1). Old contextual elements are not strengthened (re-encoded) at S2, nor are they integrated with new elements at S2. Repetition results in better memory to the extent that a greater number of contextual elements are associated with a given item. Prominent memory theorists have posited that the timing of a repetition must influence its degree of encoding variability (e.g., Martin, 1968; Melton, 1967, 1970; Tulving, 1968). Distributed repetitions are more likely to be associated with contextual elements that are different from each other (i.e., more opportunity for context to fluctuate) compared to massed repetitions, where contextual elements are likely to be the same or very similar. Therefore, distributed repetitions will be associated with a greater number of unique contextual cues than massed repetitions. Following Tulving and Thompson's (1973) encoding specificity principle, the greater the number of contextual elements associated with an item during study, the higher the probability that one or more of these elements will overlap with a cue at test, thereby assisting item retrieval.

Glenberg (1979) presented his components-level theory as the first in-depth application of encoding variability principles to the distributed practice effect in verbal learning. He proposed that an item is represented in memory as an episodic trace with multiple components including: contextual, structural, and descriptive. Contextual components represent the context in which the item is presented, akin to the concept of contextual fluctuation in stimulus sampling theories. Structural components represent the structure that a learner imposes during study, such as when a to-be-learned item is associated with another to-be-learned item(s) using a mnemonic strategy (e.g., neighbouring items on a list or semantic relatedness among items). Descriptive components represent the unique features of a given to-be-learned item. In the case of verbal learning, these features include orthography, articulation, and meaning, which together form the item's representation in lexical memory. The three components can be visualized as a hierarchy from general (most likely related to all items; contextual) to less general (most likely related to some items; structural) to specific (most likely related to only one item; descriptive). Like his predecessors, Glenberg stipulated that the episodic trace of an item repeated after a distributed lag is more likely to include a greater number of contextual and/or structural components than the trace of a stimulus repeated after a massed lag, in turn increasing the probability of its successful retrieval at test (i.e., trace/cue overlap). He further proposed that trace activation at test is inversely related to the generality of the item's components during study, meaning that specific components will enhance the probability of retrieving the item at test compared to general components.

As a test of the components-level theory, Glenberg (1979; Exp 1) asked participants to learn semantically related word pairs (*stimulus-response*) at massed (zero) and distributed lags (varying from two to six intervening items). For some repeated items, the stimulus and response terms were the same at S1 and S2 (e.g., *spoon-knife*; *spoon-knife*); for other repeated items, the stimulus term changed at S2 to a different semantically related word (e.g., *spoon-knife*; *blade-knife*). Glenberg reasoned that this paradigm empirically manipulated contextual components (i.e., massed vs. distributed lags) and descriptive components (i.e., same vs. different stimulus word) during study (he did not place constraints on how participants attempted to learn the items; therefore, he did not explicitly manipulate structural components). On a free recall test of the response items (i.e., *knife*), participants recalled more distributed items than massed items, both when encoding remained consistent between S1 and S2 and when encoding varied between S1 and S2. Glenberg reasoned that, without an explicit descriptive cue at test, the most recallable items will be those characterized by the greatest number of retrieval routes, in this case, items with varied contextual elements (i.e., the effect of lag), varied structural elements (i.e., the effect of encoding strategy), and varied descriptive elements (i.e., the effect of two referent words during study). Appropriately, the best recalled items in this experiment were those studied under distributed, variable encoding conditions[2].

---

[2] Components-level theory predicts a different outcome when memory is tested by cued recall. In cued recall, participants are provided with a descriptive cue at test that should (theoretically) exactly match information stored as a descriptive element in the item's memory trace (e.g., *spoon* or *blade*). Because descriptive components are the most specific of all components, they will be most strongly activated at test, diminishing the influence of other, more general components (in this case, contextual components; i.e., the effect of lag). Therefore, variability in descriptive elements at S1 and S2 will induce maximum encoding variability of all items in a cued recall paradigm, with little to no influence from the passing of time. The result is that under *variable* encoding conditions,

Unlike traditional encoding variability theories, which posit that repetitions are encoded as two independent memory traces, components-level theory endorses trace dependency. Glenberg (1979) acknowledged work from Johnston and Uhl (1976) and Madigan (1969) showing that distributed practice effects were only apparent for items recognized as repetitions at S2. These findings suggest that the benefit of a second study opportunity is only useful to the extent that the first study opportunity is remembered; in other words, S2 is dependent on S1. Components-level theory takes the position that S1 and S2 traces are not independent from each other. Encoding at S1 and S2 can be represented as a single, elaborate memory trace for a to-be-learned item so long as descriptive components remain functionally the same (i.e., the item itself can be recognized as a repeat). Trace dependency has received support from other research showing that memory for repeated items (one item, twice presented) is better than memory for two unique items separated by the same lag (two items, once presented), implying that repeated items cannot be analogous to two unique memory traces (Ross & Landauer, 1978).

Study-phase retrieval theory. The concept of trace dependency inspired a final class of distributed practice theories known as study-phase retrieval (e.g., Greene, 1989; Thios & D'Agostino, 1976). Study-phase retrieval theories explain the distributed practice effect based on the retrieval success of an item's earlier presentation when it is

---

in a *cued recall* scenario, massed and distributed items are more likely to be remembered at a similar (usually high) rate. However, under *constant* encoding conditions, in a *cued recall* scenario, massed and distributed items offer the same probability of trace-cue matching for descriptive components. In this case, contextual elements, gathered from the passing of time during study, are more likely to influence the cued retrieval process, in turn, revealing a distributed practice effect. Data to support these claims is also presented in Glenberg's experiment.

represented for study at S2. A core assumption of study-phase retrieval is that repetition strengthens a previously encoded memory trace rather than creates a new memory trace. When an item is represented at S2, the learner is reminded of its presentation at S1, spontaneously retrieving the item's S1 memory trace and updating it with new information encoded at S2 (Benjamin & Tullis, 2010; Hintzman, 2004). The benefit of this retrieval process on later memory is inversely related to the item's accessibility at S2 [Melton's (1970) "strength paradox" or Schmidt & Bjork's (1992) "desirable difficulty"]. The more difficult the retrieval at S2 (i.e., the less accessible the item), the more the memory trace is strengthened, assuming retrieval is successful. For massed items, retrieval of S1 at S2 will be relatively easy for the learner, having just been exposed to the item. For distributed items, retrieval of S1 at S2 is more likely to be effortful, whether due to partial forgetting and/or poor retrieval cues at S2. If retrieval of S1 is completely unsuccessful at S2, however, there should be no advantage of distributed practice on subsequent memory.

In an experiment testing principles of study-phase retrieval theory (Thios & D'Agostino, 1976), participants read aloud a series of sentences at S1 each containing a subject noun and an underlined object phrase (e.g., *The conductor boarded the <u>express train</u>*). The sentences reappeared at lags of 0, 4, or 12 intervening sentences. Some of the sentences re-appeared at S2 exactly as they had appeared at S1 and participants were then instructed to orally state the sentences' passive transformations (i.e., "*The express train was boarded by the conductor*"). Other sentences reappeared at S2 as only the object phrase (i.e., *<u>express train</u>*). Participants were also asked to come up with the passive transformations of the complete original sentences but on their own, using only the object phrase cue. Thus, the authors manipulated both lag as well as whether an item was forcibly retrieved at S2. Free recall data of the object phrases revealed a distributed practice effect only for items forcibly retrieved at S2. The probability of successful retrieval at S2 was perfect for massed items, .60 for lag 4 items, and .59 for lag 12 items. This latter finding verifies that retrieval at S2 was more difficult for distributed items, which the authors reasoned was the mechanism that improved subsequent retrieval at test.

A hybrid account. Deficient processing, encoding variability, and study-phase retrieval theories each come to different conclusions about what key mechanism drives the distributed practice benefit. It may be the result of superior reprocessing of an item, a greater number of unique contextual elements associated with an item from the drift of time, or the paradoxical influence of retrieval difficulty during restudy. Another possibility, initially proposed by Greene (1989), is that multiple mechanisms work in concert to produce the distributed practice effect. In the contemporary hybrid account, massed items are hypothesized to undergo deficient processing at S2 whereas distributed items are hypothesized to benefit from a balance of encoding variability and study-phase retrieval effort at S2 (Karpicke, Lehman, & Aue, 2014; Mozer, Pashler, Cepeda, Lindsey, & Vul, 2009; Raaijmakers, 2003; Siegel & Kahana, 2014). Evidence in support of the hybrid account comes from research showing that when a number of lags are tested in a single experiment, the optimal lag is one that is spaced (i.e., benefiting from encoding variability), but not spaced to the detriment of successful study-phase retrieval. For example, Verkoeijen, Rikers, and Schmidt (2005) repeated words at six different lags (0, 2, 5, 8, 14, and 20 intervening items) and found that participants' immediate free recall performance was best for lag 14 items, with an observable drop in performance for lag 20 items. Although the authors did not explicitly measure the effort/success of study-phase retrieval at S2, their data support the logic of a trade-off between the benefits of encoding variability and the success of study phase retrieval for distributed items. The authors reasoned that lag 14 items maintained the balance of encoding variability and successful study phase retrieval at S2 while lag 20 items, although advantaged by encoding variability mechanisms, did not undergo successful study-phase retrieval at S2 and hence were not recalled as well as lag 14 items on the final test. Mathematical models provide additional support for the hybrid account by successfully predicting final recall data using forgetting curves across multiple lags (e.g., Mozer et al., 2009; Raaijmakers, 2003).

Factors Influencing the Distributed Practice Effect

The distributed practice effect is famously robust; however, there are at least two factors that reliably influence its magnitude. These factors include the timing of the retention interval and the consistency of the study context.

The Lag × Retention Interval interaction. Research shows that the advantage of distributed practice is reduced when final memory is tested at very short retention intervals. This finding, referred to as the Lag × Retention Interval interaction, was first discussed by Peterson and colleagues (1963) and later formalized by Glenberg (1976). In Glenberg's study, participants learned word pairs repeated at lags of 0, 1, 4, 8, 20, or 40 intervening items that were later tested (cued recall) after retention intervals of 2, 8, 32, or 64 intervening items. When tested at the long retention interval of 64 items, participants remembered more items repeated at the long lag (lag of 40); however, when tested at the short retention interval of 2 items, participants remembered more items repeated at a shorter lag (lag of 4). The Lag × Retention Interval interaction implies that spaced repetition does not always improve subsequent memory; rather, the optimal time to relearn a given set of material depends on the retention interval. It has similarly been reported for between-session paradigms, which corroborates popular student opinion that “cramming” the night before an exam can be an effective study strategy (Cepeda et al., 2008; Pyc, Balota, McDermott, Tully, & Roediger, 2014; Rawson & Kintsch, 2005).

The effect of the retention interval on the distributed practice effect can be accounted for using the hybrid account, with an emphasis on encoding variability mechanisms. Glenberg (1976, 1979) reasoned that the best study schedule is one in which encoding contexts during study and test have the greatest overlap (Tulving & Thompson, 1973). Massed study is assumed to share the most contextual overlap with the test environment if the test occurs immediately (when test cues are recent and predictable). Therefore, repetition at a short lag will be optimal for a short retention interval but not a long retention interval. Distributed study is assumed to share the most contextual overlap with the test environment if the test occurs in the future (when test cues are unpredictable). Therefore, repetition at a long lag will be optimal for a long retention interval but not a short retention interval.

The Lag × Study Context interaction. Research also shows that the advantage of distributed practice is reduced when the second presentation of an item does not exactly match its first presentation, for example, when the experimenter changes features associated with a to-be-learned item (i.e., study context) throughout the study phase. This finding, referred to as the Lag × Study Context interaction, is of particular relevance to the present dissertation and will be discussed in greater detail. Considering that real-world learning is often characterized by unpredictable study contexts, the replicability of this finding, its implications, and its connection to distributed practice theories are of great importance.

A targeted search of the literature reveals 35 single-session distributed practice studies where experimenters explicitly manipulated study context at S1 and S2, measured participants' final cued recall, free recall, and/or recognition memory, and tested the Lag × Study Context interaction (Table 1). Figure 2 depicts exemplar data illustrating a distributed practice effect for items repeated in a consistent study context during the study phase (e.g., the word *apple* presented on a blue background at S1 and S2) but no distributed practice effect for items repeated in a variable study context during the study phase (e.g., the word *banana* presented on a blue background at S1 and a red background at S2). The interaction is driven by the positive influence of variable study on later memory for massed items (i.e., massed, variably studied > massed, consistently studied) and no influence (sometimes a detrimental influence) on later memory for distributed items (i.e., distributed, variably studied = distributed, consistently studied; exceptions are noted in Table 1). The Lag × Study Context interaction implies that spaced repetition does not always improve subsequent memory; rather, the optimal time to relearn a given set of material depends on the manner in which it is restudied.

The Lag × Study Context interaction has been examined using a variety of study context manipulations that depend on the type of to-be-learned stimuli. In the case of words and nonwords, items have been restudied using either the same or different: modality (e.g., auditory vs. visual; Wells & Kirsner, 1974), orienting task (e.g., perceptual vs. semantic; Bird et al., 1978), rating scale (e.g., imageability vs. pleasantness; Greene & Stillwell, 1995), language (e.g., bilingual participants studying in the same language vs. in different languages; Glanzer & Duarte, 1971), background scene (e.g., same colour vs. different colour; Verkoeijen, Rikers, & Schmidt, 2004), or typography (e.g., same font vs. different font; Mammarella, Avons, & Russo, 2004). In the case of word pairs, items have been restudied using either the same or different: stimulus words (e.g., *speed-engine* vs. *valve-engine*; Madigan, 1969) or sentence frames (e.g., The high powered *drill* entered the masonry *blocks* vs. The fire *drill* cleared the city's *blocks*; Thios, 1972). Several word pair paradigms have capitalized on the lexical nature of homographs, which are words that evoke two different semantic interpretations depending on the stimulus word with which they are presented (e.g., *flower-bulb* vs. *light-bulb*; Hintzman, Summers, & Block, 1975). In the case of sentences and text passages, items have been restudied either verbatim or in paraphrased form (e.g., Dellarosa & Bourne, 1985; Glover & Corkhill, 1987). Finally, in the case of non-verbal learning, objects have been restudied using either the same or different: modality (e.g., names of objects vs. pictures of objects; Paivio, 1974), background scene (e.g., von Wright, 1976, Exp 1), or item exemplar (e.g., von Wright, 1976, Exp 2). Faces have been restudied using either the same pose or a different pose (e.g., headshot vs. side profile shot; Mammarella, Russo, & Avons, 2002).

The effect of study context on the distributed practice effect can be accounted for using the hybrid account. First, a deficient processing mechanism assumes that massed, variably studied items are less redundant than massed, consistently studied items. Thus, massed, variably studied items are more likely to undergo full reprocessing at S2, improving their chances of future retrieval relative to massed, consistently studied items, which undergo deficient processing at S2. Second, encoding variability and study-phase

Table 1

Summary of Studies Investigating the Lag × Study Context Interaction Effect

Author(s)	Year	Exp(s) of Interest	Stimuli	Study Context Manipulation	Test Format	Sig. Lag × Study Context Interaction
Madigan	1969	2	Word pairs	Cue word	Free recall / Cued recall	✓
Gartman & Johnson	1972	3	Word pairs	Cue word	Free recall	✓
Shaughnessy et al.	1974		Word pairs	Cue word	Free recall / Cued recall	
Glenberg	1979	1	Word pairs	Cue word	Free recall / Cued recall	
Belleza & Young	1989	3	Word pairs	Cue word	Cued recall	✓
Johnson et al.	1972		Word pairs (Homographs)	Cue word	Free recall	
Hintzman et al.	1975	1 and 2	Words / Word Pairs (Homographs)	Cue word	Recognition	✓
Winograd & Raines	1972		Word pairs (Homographs)	Semantic properties	Recognition	✓
Glanzer & Duarte	1971		Words	Language	Free recall	✓

Author(s)	Year	Exp(s) of Interest	Stimuli	Study Context Manipulation	Test Format	Sig. Lag × Study Context Interaction
Paivio et al.	1968	1 and 2	Words	Language	Free recall	✓
Verkoeijen	2005	4 ABC	Words	Language	Free recall	✓
Wells & Kirsner	1974		Words	Modality of presentation	Free recall	
Glenberg & Smith	1981	1 and 2	Words	Modality of presentation / Orienting task	Free recall / Recognition	✓
Maskarinec & Thompson	1976	1 and 2	Words	Orienting task	Free recall	
Shaughnessy	1976	2 and 3	Words	Orienting task	Free recall	
Bird et al.	1978	1 and 2	Words	Orienting task	Free recall	
Jensen & Freund	1981	1 and 2	Words	Orienting task	Free recall	
Toppino & DeMesquita	1984	2	Words	Orienting task	Free recall	✓

Author(s)	Year	Exp(s) of Interest	Stimuli	Study Context Manipulation	Test Format	Sig. Lag × Study Context Interaction
McFarland et al.	1979	1 and 2	Words	Orienting task / Semantic properties	Free recall	✓
Rose	1980	1 and 2	Words	Semantic properties	Recognition	✓
Rose	1984	1 and 2	Words	Semantic properties	Free recall / Recognition	✓
Verkoeijen et al.	2004	1 and 2	Words	Background	Free recall	✓
Hockley et al.	2012	3	Words	Background	Free recall	
Russo et al.	2002	1, 2, and 3	Nonwords	Font	Recognition	✓
Mammarella et al.	2004	1 and 2	Nonwords	Font	Recognition	✓
Thios	1972		Sentences	Semantic properties	Cued recall	✓
D’Agostino & DeRemer	1973	1 and 2	Sentences	Semantic properties	Free recall / Cued recall	✓
D’Agostino	1974	1 and 2	Sentences	Semantic properties	Free recall	✓

Author(s)	**Year**	**Exp(s) of Interest**	**Stimuli**	**Study Context Manipulation**	**Test Format**	**Sig. Lag × Study Context Interaction**
Dellarosa & Bourne	1985	1 and 2	Sentences	Phrasing / Modality of presentation	Free recall	✓
Glover & Corkhill	1987	1 and 2	Text passages	Phrasing	Free recall	✓
Durgunoglu et al.	1993	2	Text passages	Language	Comprehension	✓
Paivio	1974	1 and 2	Objects	Modality of presentation	Free recall	✓
von Wright	1976	1 and 2	Objects	Background / Item exemplar	Free recall	✓
Mammarella et al.	2002	1, 2, and 4	Faces	Pose	Recognition	✓
Appleton-Knapp et al.	2005	2 and 4	Advertisements	Appearance of ad / Modality of presentation	Cued recall / Recognition	✓

Note. Studies are organized by type of stimuli and study context manipulation. For conciseness, studies using a frequency discrimination test are excluded.

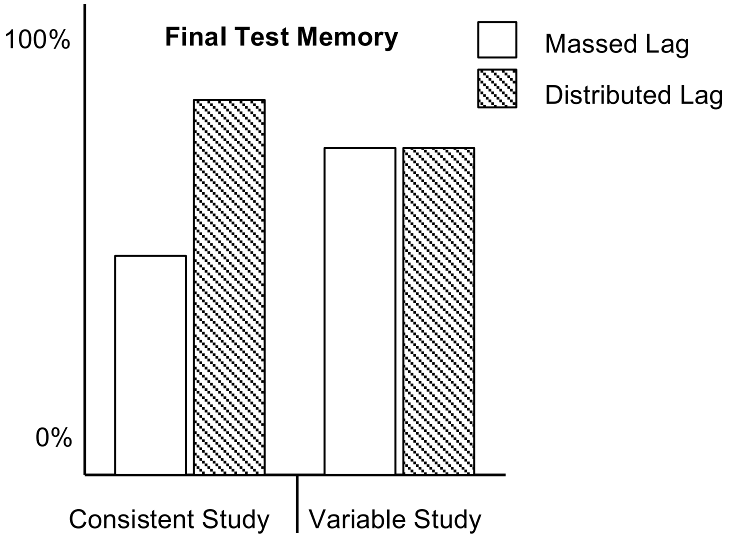

Figure 2. Hypothetical data illustrating a Lag × Study Context interaction. *Consistent study* refers to item repetition in the same study context throughout the study phase, for example, when the word *apple* is studied on a blue background at S1 and again on a blue background at S2. *Variable study* refers to item repetition in two different study contexts throughout the study phase, for example, when the word *banana* is studied on a blue background at S1 and on a red background at S2. Under consistent study conditions, the distributed practice effect is robust (i.e., distributed repetition > massed repetition). However, under variable study conditions, the distributed practice effect often disappears (i.e., distributed repetition ≤ massed repetition).

retrieval mechanisms assume that distributed items benefitting from the drift of time (encoding variability) *and* effortful but successful retrieval/updating during restudy (study-phase retrieval) will have the most elaborate memory traces and therefore the best chances of recall on a final test. Distributed, consistently studied items are more likely to strike the right balance of encoding variability and effortful/successful study phase retrieval at S2. Distributed, variably studied items, although benefitting from encoding variability, are more likely to suffer from unsuccessful study phase retrieval at S2, compromising their chances of future retrieval relative to distributed, consistently studied items.

Goals of the Dissertation

The studies that make up the present dissertation provided novel contributions to understanding the effect of study context on the distributed practice effect. There were two overarching goals of the research. The first goal was to investigate whether the Lag × Study Context interaction has a neural basis and, if so, whether neural data converge with behavioural data to support cognitive processing assumptions outlined by the hybrid account. To examine these questions, Study 1 describes an experiment where electroencephalography (EEG) was continuously recorded as participants completed a cued recognition task. To-be-learned stimuli repeated after varying lags on either the same background colour or a different background colour. After a fixed (within-list) retention interval, stimuli were presented for a third time and participants' recognition memory was tested. Event-related potentials (ERPs) collected during study and test trials were compared across repetition conditions. This experiment also served to replicate previous electrophysiological investigations of the traditional distributed practice effect without the study context factor (e.g., Kim, Kim, & Kwon, 2001; Mollison, 2015; Van Strien, Verkoeijen, Van der Meer, & Franken, 2007).

The second goal of the research was to extend previous behavioural investigations of the Lag × Study Context interaction to non-undergraduate samples. The interaction has been replicated several times with undergraduates; however, its generalizability to other populations—in particular, older adults—has not been explored. A large body of evidence suggests that older adults encode and retrieve information about an item's study context less successfully than younger adults (e.g., Cansino et al., 2013; Old & Naveh-Benjamin, 2008; Spencer & Raz, 1995); therefore, study context may have a different influence on the distributed practice effect for this age group. To explore this question, Study 2 describes two experiments where undergraduate participants and older adult participants aged 60+ years learned stimuli repeated after varying lags on either the same background scene or a different background scene. The background scenes were either shared among all to-be-learned items (Experiment 2A) or unique to each to-be-learned item (Experiment 2B). After the study phase, participants' free recall memory was tested. Data were examined in light of a hypothesized Age × Lag × Study Context interaction.

Study 1

Compared to the behavioural literature, relatively few studies have investigated the neural basis of the distributed practice effect (e.g., Van Strien et al., 2007; Xue et al., 2011). Neuroimaging is a complementary tool for exploring potential cognitive processing differences between massed and distributed repetitions. While behavioural data are often limited to a measure of participant response after a stimulus has been presented [e.g., reaction time (RTs) or accuracy], neural data offer insight into brain mechanisms that are likely operating between stimulus and response, independent of a participant's overt behaviour. Neural data may also be used to speculate about the cognitive processes involved in a given behavioural phenomenon, thereby informing theory development. In Study 1, neural mechanisms of the Lag × Study Context interaction were explored. ERPs were recorded as participants studied items repeated at a massed lag or a distributed lag that were subsequently tested for recognition after a fixed retention interval. The influence of study context on the neural representation of the distributed practice effect was also examined by manipulating the background colour on which an item appeared. This study is the first to measure ERPs in a complete distributed practice paradigm (i.e., S1, S2, and final test) with the goal of understanding the effect of study context on the distributed practice effect from a neural perspective.

Understanding Event-Related Potentials (ERPs)

Electroencephalography (EEG) is a neuroimaging technique that is widely used in research on the human brain. As thoroughly described in a reference text by Luck (2005), EEG measures the electrical activity of neurons in the cerebral cortex using non-invasive electrodes applied to the surface of the scalp. EEG waveforms are characterized by their frequency (i.e., number of cycles per second, measured in Hz), polarity (i.e., positive or negative, relative to a reference point), amplitude (i.e., mean/peak voltage, usually measured in microvolts, μV), and scalp distribution (i.e., region of peak electrodes). ERPs are simply EEG waveforms that have been time-locked to a specific internal or external event and averaged across many trials. Thus, in addition to the characteristics noted above, ERPs can also be described by their latency (i.e., time to reach peak voltage after stimulus onset, measured in ms). Researchers typically design ERP studies to examine one or more ERP component(s), which can be defined as "scalp-recorded voltage changes that reflect specific neural or psychological processes" (Kappenman & Luck, 2012, p. 4). The accumulation of converging evidence across different experimental paradigms has allowed researchers to confidently draw correlational associations between specific ERP components and neural or psychological processes. It is customary to label ERP components according to their polarity (positive/P or negative/N) and approximate latency (either abbreviated or long-form). For example, as described in the next section, evidence suggests that the N1 component is associated with attentional processing (negative peak, onsetting ~100 ms post-stimulus) and the later N400 is associated with semantic processing (negative peak, onsetting ~400 ms post-stimulus). Latency values will vary across studies; therefore, it is standard practice to examine electrical activity during a defined time window of interest rather than at one specific time point (e.g., 300-500 ms for the N400).

Understanding what exactly is being measured by ERPs requires an understanding of how neurons in the brain communicate with one another. When a neurotransmitter binds to receptors on a cell's membrane, the flow of ions across the membrane changes, causing the electrical transmission of either an inhibitory or excitatory signal that travels down the cell's axon. This process is known as a post-synaptic potential (PSP). A PSP creates an electrical dipole: an oriented flow of current where one end of the neuron is more positively charged and the other end is more negatively charged (hence the potential for electrical energy to flow across the dipole). ERPs capture the summed activation of PSPs occurring simultaneously across similarly oriented neurons in the cortex. Importantly, the summation must be large enough (i.e., the contributions of thousands of neurons) to pass through brain tissue, meninges, and skull in order to be captured by electrodes placed on the scalp. In addition, the electrodes can only capture the summed electrical activity of neurons oriented in the same direction otherwise the dipoles of neighbouring neurons will cancel each other out and, theoretically, electrical activity recorded at the scalp will equal zero. Of all the different types of cells in the human brain, ERPs most likely reflect the synchronized post-synaptic activity of pyramidal cells since they are oriented perpendicular to the cortical surface and are typically in phase with their neighbours. Despite imperfect spatial resolution, ERPs have unparalleled temporal resolution, allowing researchers to examine exactly when the brain is responding to a given stimulus as well as the sequence of processes that likely precede a motor response initiated by the participant. Informative data can also be obtained from ERPs in the absence of an overt response, which is a methodological advantage of EEG technology that has revolutionized many areas of cognition research. For example, in studies of attention, researchers have used ERPs to compare neural responses to attended versus ignored stimuli, the latter which demand no participant response and are therefore difficult to study using traditional behavioural measures.

ERP Components of Interest

In order to evaluate the neural basis of the distributed practice effect in the present study, ERP components relevant to distributed practice theories were considered. Neural measures were selected by reviewing the broad literatures of ERPs associated with attention and memory (Friedman & Johnson, 2000; Luck, Woodman, & Vogel, 2000; Rugg & Curran, 2007) as well as more specific literature on ERPs associated with word repetition across different lags (Kim et al., 2001; Mollison, 2015; Van Strien et al., 2007). Three ERP components were common across the range of studies reviewed: the visual N1, FN400, and late positivity component (LPC). Each of these components has an important connection to theories of the distributed practice effect.

ERP studies of attention and memory.

N1 and visual attention. The N1 wave is an early sensory-evoked potential that has been associated with featural analysis in the visual domain (Luck et al., 2000). It may have important connections to deficient processing theories of the distributed practice effect (Mollison, 2015). In one pivotal study, Vogel and Luck (2000) tested the hypothesis that the N1 reflects attention processes associated with task-general visual discrimination. Participants responded to five-letter, multi-coloured stimulus arrays either by pressing a single button on every single trial (simple RT: no discrimination) or by pressing one of two buttons to indicate whether the trial contained a particular letter (choice RT: form discrimination) or colour (choice RT: colour discrimination). Supporting their hypothesis, trials of the two choice RT conditions, which did not differ from each other in N1 mean amplitudes, produced larger N1 mean amplitudes compared to trials of the simple RT condition. The authors reasoned that the N1 signified top-down featural analysis of a stimulus. Further analyses demonstrated that the N1 component could be dissociated into two subcomponents, which differed in their topographies and relations to motor processing. The first N1 subcomponent, peaking ~100-150 ms post-stimulus at frontocentral electrode sites, shared significant overlap with motor ERPs known to be associated with choice RT tasks. The second N1 subcomponent, peaking ~150-200 ms post-stimulus at posterior occipital sites, was preferentially engaged during task discrimination without the motor control confound. A second experiment using simple versus choice RT tasks that did not include a motor response confirmed that there were no mean amplitude differences across conditions for the early N1 subcomponent; however, mean amplitude differences remained between simple and choice RT tasks for the later N1 subcomponent. A third experiment revealed that the later N1 subcomponent was insensitive to perceptual demands of the discrimination task [i.e., the waveforms were similar for tasks using alike colours/forms (hard to discriminate) versus different colours/forms (easy to discriminate)]. Vogel and Luck's experiments provide convincing evidence that the posterior N1 component, hereafter referred to as simply the visual N1, is involved in early attention to the visual features of a stimulus regardless of motor response or perceptual load.

Since the visual N1 reflects stimulus-evoked attention, it may be an informative neural correlate of deficient processing mechanisms that have been used to explain the distributed practice effect. Deficient processing theory suggests that massed items are poorly reprocessed at S2. On the contrary, distributed items receive full reprocessing at S2, which improves their subsequent retrieval at final test. Others have proposed that attenuation of the visual N1 component is one way of quantifying the neural basis of deficient processing for massed repetitions relative to distributed repetitions (Mollison, 2015). Previous studies associating visual N1 attenuation with perceptual priming and word priming also support this proposal (e.g., Posner & McCandliss, 1999; Wiggs & Martin, 1998).

FN400, LPC, and recognition memory. Electrophysiology studies have made important contributions to our understanding of memory (e.g., Curran, Tepe, & Piatt, 2006; Friedman & Johnson, 2000; Rugg & Allan, 2000; Rugg & Curran, 2007). For example, since the 1980s, it has been well documented that ERPs evoked by repeated (old) items are more positive-going than ERPs to once-presented (new) items, a shift which occurs ~200 ms post stimulus and may persist for seconds across a range of electrode sites but especially central sites (e.g., Rugg, 1995, Figure 5). This ERP "old/new repetition effect" provides a foundation for the neural basis of recognition memory.

Beyond the old/new repetition effect, ERPs have been used to elucidate whether recognition memory is best understood from the framework of a single- or dual-process model (e.g., Dunn, 2004). A single-process model likens the process of recognition to signal detection theory, where positive memory judgments are made when the strength of evidence exceeds a criterion level. A dual-process model posits that there are two functionally distinct memory signals that underlie recognition, one of which is based on a strength-of-signal interpretation (familiarity) and the other which is based on the retrieval of episodic information related to the study event (recollection). Evidence from ERPs consistently supports the dual-process model, revealing two quantitatively and qualitatively different neural responses that each reflect unique retrieval-related processing: (1) an early (~300-500 ms) positivity maximal over midline frontal sites, considered to be a neural correlate of familiarity and (2) a later and longer-lasting (~400-1000 ms) positivity maximal over left/central parietal sites, considered to be a neural correlate of recollection. As explained below, these dissociable components may have important connections to encoding variability and study-phase retrieval theories of the distributed practice effect.

A neural distinction between familiarity and recollection retrieval processing receives support from a number of different experimental paradigms, including: lexical decision, plurality recognition, and levels of processing. In a lexical decision paradigm, participants classify stimuli as words or nonwords. Curran (1999) measured ERPs during lexical decision followed by a subsequent recognition test. ERPs to correctly recognized old items were characterized by two distinct components: an early frontal old/new effect that was similar in mean amplitudes for words and nonwords combined with a later parietal old/new effect that was larger in mean amplitude for words compared to nonwords. According to the author's rationale, participants likely recognized words as being both familiar and recollected, therefore engaging both a frontal old/new effect and a parietal old/new effect for this class of stimuli (e.g., "The word *apple* looks familiar and I also recollect the mental image of eating an *apple*"). However, in the absence of semantic properties, participants likely recognized nonwords only on the basis of familiarity, therefore engaging a weaker parietal old/new effect for nonwords compared to words (e.g., "The nonword *cherk* looks familiar"). Rugg and Doyle (1992) used the same logic and further reported that when word frequency was added as a factor in the lexical decision paradigm, the parietal old/new effect was larger in mean amplitude for low frequency words compared to high frequency words, especially those that participants correctly classified as old with high confidence. The authors reasoned that correctly recognized, low frequency/high confidence stimuli were most likely to engage recollection-like retrieval compared to all other stimuli.

Plurality recognition has also been used as a means of differentiating familiarity and recollection retrieval processing at the neural level. In this paradigm, participants study a list of singular and plural nouns (e.g., *truck*, *plants*) and then complete a recognition test comprising distractors that are distinct from targets (e.g., *ribbon*? *pencils?*) or similar to targets but opposite in plurality (e.g., *trucks? plant?*) It is proposed that a false alarm to a similar distractor signifies familiarity in the absence of recollection. Curran (2000) recorded ERPs as participants completed a plurality recognition paradigm. ERPs to correctly recognized targets and falsely recognized similar distractors engendered frontal old/new effects that were of similar mean amplitudes (i.e., both classes of stimuli were familiar) and a parietal old/new effect that was larger in mean amplitude for correctly recognized targets (i.e., familiarity with recollection) compared to falsely recognized similar distractors (i.e., familiarity without recollection).

Finally, levels of processing manipulations have confirmed neural differences between familiarity and recollection retrieval. In a levels of processing paradigm, participants are instructed to study word stimuli using shallow (e.g., orthographic) or deep (e.g., semantic) encoding strategies. Rugg and colleagues (1998) collected ERPs as participants completed a levels of processing paradigm followed by a subsequent recognition test. The ERP data revealed frontal old/new effects that were of similar mean amplitudes for words studied in both encoding conditions combined with a parietal old/new effect that was largest in mean amplitude for words studied in the deep encoding condition. Similar findings were reported in an ERP source memory study (Wilding & Rugg, 1996) where participants studied words presented aurally by either a male or female speaker that were later tested for item (old/new) and source (male/female speaker) recognition. Frontal old/new effects were similar in mean amplitudes for hit/hit targets and hit/miss targets; however, the parietal old/new effect was largest in mean amplitude and took longer to peak for hit/hit targets.

The selection of studies reviewed above demonstrates the range of ERP evidence supporting a dual-process model of recognition, where two quantitatively and qualitatively different memory signals (i.e., signals differing in amplitude and/or latency and scalp topography) distinguish familiarity-like recognition from recollection-like recognition. Familiarity processing is observed first in the form of a greater positivity ~300-500 ms post-stimulus at anterior sites. Familiarity implies that an item engages a threshold level of remembering; however, it does not engage reinstatement of a prior study experience. Instead, such episodic details are reflected in later recollection processing, which is observed as a slow-going positivity ~400-1000 ms post-stimulus at posterior sites.

Aside from the general nomenclature “frontal old/new effect” and “parietal old/new effect”, familiarity and recollection ERP components are also referenced in the literature under the labels of FN400 and late positivity component (LPC), respectively. To be clear, the FN400 is assumed to share functional characteristics with the N400 component described in ERP studies of language processing. In these studies, the N400 is widely recognized for its role in semantic processing (for a review, see Kutas &

Federmeier, 2011). It is maximal (i.e., most negative) over centroparietal sites when an item is evaluated for meaning, especially for cases where there is semantic incongruence (e.g., "I take my coffee with cream and *frog*"). The N400 is reliably attenuated by word repetition; however, the effects are maximal over frontocentral sites (Friedman & Johnson, 2000). To differentiate the N400 of language processing (parietally distributed) from the N400 repetition effect (frontally distributed), Curran (1999) adopted the term frontal N400 or FN400. When the FN400 component is attenuated (i.e., more positive, as in the studies reviewed above), it is said to reflect less semantic processing because of a greater degree of familiarity with the item.

Since the FN400 and LPC components have been linked to recognition processing, they may be informative neural correlates of encoding variability and study-phase retrieval mechanisms, respectively, each of which has been used to explain the distributed practice effect. Encoding variability theory states that distributed repetitions are associated with a greater number of unique contextual cues encoded during study compared to massed repetitions. Since retrieval success on a final test is gauged by the degree of overlap between study cues and test cues, distributed repetitions (many study cues) are more likely to prevail over massed repetitions (few study cues). Attenuation of the FN400 on the basis of item familiarity relatedly implies the absence of any new encoding activity (i.e., if the item is familiar it need not be re-encoded via semantic processing). From the reverse perspective, little/no attenuation of the FN400 relatedly implies the presence of new encoding activity (i.e., if the item is not familiar it must be re-encoded via semantic processing). Therefore, the FN400 component is one way of quantifying the neural basis of encoding variability for spaced repetitions relative to massed repetitions.

Study-phase retrieval theory explains the distributed practice effect by assuming that, upon repetition of an item, the item's initial presentation is retrieved and its memory trace updated. Critically, it is assumed to be more effortful to retrieve the initial presentation of an item if its repetition occurs after a space of time compared to immediately when the item may still be active in working memory. Retrieval of S1 at S2 is more difficult and/or may take longer for distributed repetitions compared to massed repetitions, but ultimately the attempt, when successful, serves to strengthen and update the memory trace for final retrieval. The LPC component, which is observed as a larger amplitude and/or longer latency for recollection-like retrieval processing, is one way of quantifying the neural basis of study-phase retrieval for spaced repetitions relative to massed repetitions.

In summary, based on a survey of the relevant literatures, the ERP components of interest in the present study include the visual N1, FN400, and LPC. Broadly defined, these components are known for their roles in attention (deficient processing theory), familiarity (encoding variability theory), and recollection processing (study-phase retrieval theory), respectively. These components have also been the foci of previous ERP studies of word repetition across different lags.

ERP studies of word repetition across different lags. Early applications of ERP to the study of memory demonstrated that the old/new repetition effect was realized differently depending on the specific timing of a repetition (e.g., Friedman, 1990; Nagy & Rugg, 1989; Segalowitz, Van Roon, & Dywan, 1997). Recent research has further specified differences in amplitude and/or latency of the visual N1, FN400, and/or LPC components when repetitions across different lags are compared.

In a study by Kim and colleagues (2001), participants completed a continuous recognition task to a series of Korean words. Words were presented either once or twice at lags of either 0 or 5 intervening items. ERPs to massed items elicited an attenuated FN400 and an earlier-to-peak LPC compared to distributed items. Conversely, ERPs to distributed items elicited an FN400 that was comparable in mean amplitude to new words as well as a later-to-peak LPC. The authors interpreted FN400 amplitude as the degree of semantic processing required at S2 (i.e., massed < distributed = new). Differences in LPC latencies between the two repetition conditions were summarized as ease of contextual re-enactment at S2 (i.e., distributed > massed), an interpretation which was corroborated by longer RTs to distributed items compared to massed items. This study did not include a retention interval/final test; therefore, the effect of distributed practice on subsequent memory could not be confirmed. Nevertheless, the results importantly demonstrate that timing of a repetition influences whether it is re-experienced on the basis of familiarity or recollection.

Van Strien and colleagues (2007) used a nearly identical experimental paradigm as the previous study and reported converging findings. Participants completed a continuous recognition task to a series of Dutch words. Words were presented either once or twice at lags of either 0 or 6 intervening items. This paradigm included an unexpected (behavioural) final free recall test. In addition to observing a classic distributed practice effect in the free recall data (i.e., distributed > massed > once presented), ERPs at S2 were exactly the same as those described by Kim et al. (2001).

Finally, in experiments by Mollison (2015), participants saw word-image pairs presented once or twice at lags of either 0 or 12 intervening items (Exp 1) or 0, 2, 12, or 32 intervening items (Exp 2). ERPs were recorded at S2 and were time-locked to word onset. After the EEG recording, participants were tested (behavioural) in a final recognition/cued recall test (i.e., Is this image old/new? If old, what is the associated word?) Findings were consistent between the two experiments. A behavioural distributed practice effect was reported for both components of the test phase. The visual N1 was attenuated for massed items compared to distributed items, which was interpreted as neural support for the deficient processing theory. The [F]N400 increased in negativity with lag, which was interpreted as enhanced semantic processing of distributed items. The LPC was larger in mean amplitude but also peaked faster for massed items compared to distributed items, which was interpreted as fast and easy study-phase retrieval of massed items that were likely still active in working memory. Neural similarity between an item’s first presentation and its repetition was also assessed using time-frequency and voltage comparison analyses. The time-frequency analysis revealed that distributed repetitions were associated with greater power (i.e., synchronization) in the theta band, a frequency range known to be associated with encoding and retrieval processing. This finding was interpreted as support for encoding variability and study-phase retrieval theories of the distributed practice effect. The voltage comparison analysis showed that distributed repetitions induced a consistent representation across time (i.e., were most alike in terms of voltage at S1 and S2), whereas massed repetitions induced a more variable representation across time. While consistent with study-phase retrieval theory, this finding was considered to be in conflict with assumptions of encoding variability theory.

Goals and Hypotheses of the Current Study

By associating ERP components to all three theoretical mechanisms comprising the hybrid account, Mollison's (2015) experiments offer the most comprehensive evaluation of the neural basis of the distributed practice effect to date. One goal of the present study was to replicate these findings using a traditional distributed practice paradigm where words were presented for study either once or twice at lags of either 0 or 8 intervening items. A second goal was to address a limitation common to all three of the studies reviewed above, namely, that ERP measurements have never been collected at final test. After a sufficient retention interval, it is important to re-evaluate FN400 and LPC components for massed and distributed items as a means of comparing whether their final recognition, like their S2 recognition, is based on familiarity-like retrieval and recollection-like retrieval, respectively. To this end, the present study recorded continuous EEG during a cued recognition procedure where, after a fixed retention interval of 20 intervening items, recognition of each target item was assessed. Finally, a third goal of the present study was to extend the previous literature by comparing ERP components for items repeating at different lags in either a consistent study context or a variable study context. Research reveals that study context is an important boundary condition of the distributed practice effect; when an item is repeated in a different study context than its first presentation, the advantage of distributed practice over massed practice is often reduced (Table 1; Figure 2). The hybrid account explains the Lag × Study Context interaction using a combination of deficient processing, encoding variability, and study-phase retrieval mechanisms. Specifically, massed items are no longer expected to suffer from deficient processing under variable study conditions and, although distributed items should benefit from an even greater degree of encoding variability under variable study conditions (i.e., the drift of time *and* a change in study environment), their study-phase retrieval may be compromised. The Lag × Study Context interaction has never been investigated using neuroimaging methods. It is possible that variations to study context further distinguish neural processing differences between massed and distributed repetitions that have been reported in previous ERP research. Here, study context was operationally defined as the background colour on which an item appeared. This type of manipulation has been used in previous behavioural research on the same topic (e.g., Verkoeijen et al., 2004). Background colour was either the same at S1 and S2 (consistent study) or different at S1 and S2 (variable study). Therefore, the experiment was a 2 (Lag: 0 or 8 intervening items; within-subjects) × 2 (Study Context: consistent or variable; within-subjects) repeated measures design.

The following a priori hypotheses were made:

Hypothesis 1: A behavioural Lag × Study Context interaction was expected. Distributed targets should be recognized faster/with greater accuracy than massed targets, but only if repeated in a consistent study context at S2.

Hypothesis 2: An ERP old/new repetition effect was expected. Based on visual inspection of the ERPs, targets should be more positive-going at their S2 presentations compared to their S1 presentations (and compared to non-targets studied for the first/only time).

Hypothesis 3: An ERP Lag × Study Context interaction was expected in the visual N1 data at S2. The visual N1 should be attenuated for massed targets compared to distributed targets, but only if repeated in a consistent study context.

Hypothesis 4: An ERP main effect of Lag and Lag × Study Context interaction was expected in the FN400 data at S2. Mean amplitude of the FN400 should be greater (i.e., less attenuated) for distributed targets compared to massed targets, especially for distributed targets repeated in a variable study context.

Hypothesis 5: An ERP main effect of Lag and Lag × Study Context interaction was expected in the LPC data at S2. The LPC should be greater in mean amplitude and/or longer in latency for distributed targets, especially for distributed targets repeated in a consistent study context.

Hypothesis 6: At final test (with a focus on hits data), all targets should be more familiar (i.e., greater FN400 attenuation) than non-targets. The FN400 waveform was not expected to differ across target conditions (i.e., by a third exposure, all targets should be familiar). An ERP Lag × Study Context interaction was expected in the LPC data at final test. The LPC should be greater in mean amplitude and/or longer in latency for distributed targets compared to massed targets (and non-targets), but only if repeated in a consistent study context at S2.

Method

Participants

Prior to recruiting participants, power analyses were conducted to estimate the sample size required to detect Lag × Study Context interactions in both the behavioural and ERP data using a 95% power criterion. To detect a behavioural Lag × Study Context interaction (average *d* = 0.96, based on effect sizes taken from Appleton-Knapp, Bjork, & Wickens, 2005; Mammarella et al., 2002; Russo, Mammarella, & Avons, 2002; Verkoeijen et al., 2004), a sample size of at least 12 participants was recommended. To detect an ERP Lag × Study Context interaction (average *d* = unknown/no previous data available; therefore, based on a large effect size reported in the behavioural data), a sample size of at least 12 participants was recommended.

Twenty students were recruited from the York University Undergraduate Research Participant Pool. Individuals were eligible to participate in the experiment if they were fluent in English, had normal or corrected-to-normal vision, and they passed a colourblindess test. In exchange for participating in the experiment, they received course credit. Data from 4 participants were discarded because of technical problems during the EEG recording. Therefore, the final sample consisted of 16 participants (14 female; $M_{\text{age}} = 20.31$, $SD_{\text{age}} = 3.93$). The experiment was approved by York University's Research Ethics Board and written consent was obtained from all participants prior to beginning the session.

Materials

Colour vision test. The Pseudoisochromatic Plates Ishihara Compatible Color Vision Test was used to screen participants for red-green (RG) and yellow-blue (YB) colour perception deficiencies (Waggoner, 2005). The test has 17 plates (14 RG and 3 YB). Each plate illustrates a single- or double-digit number using dots of various sizes and colours. The participant is instructed to read aloud the number that he/she sees. The test is scored using a pass/fail method.

Word stimuli and background colours. Using the MRC Psycholinguistic Database (Coltheart, 1981), a pool of 784 words was compiled. Items were moderate-to-high frequency English nouns containing 3-7 letters and 1-3 syllables. They were concrete and imageable (ratings > 400, on a scale ranging from 0-1000).

Items were randomly assigned to one of four lists (corresponding to four blocks of the experiment) and further randomly assigned to serve as targets in one of the four experimental conditions or to serve as non-targets. In each list, each condition was represented by 20 target words that were studied twice and tested once. The total number of tested targets (20*4 = 80) was complemented by the same number of tested non-targets (80), which were words that were never studied but tested once. In addition, there were 20 studied non-target items (studied once, never tested), 8 primacy buffers (studied at the start of the list, never tested), and 8 recency buffers (studied at the end of the list, never tested). Therefore, the total number of unique words per block was 196 and the grand total of unique words across the entire experiment was 784 (196*4). Orange and green were selected as the background colours on which items appeared during study.

Stimulus presentation. To ensure that targets from the four experimental conditions were evenly presented throughout a list, a custom programming script was written in MATLAB. The script began as an empty array consisting of 356 positions[3]. First, positions for the primacy and recency buffers were assigned. Second, the S2 presentation of each target in each condition was distributed evenly throughout the list

---

[3] Per experimental condition, there were 20 targets that each appeared three times. There were also 80 tested non-targets that each appeared once, 20 studied non-targets that each appeared once, 8 primacy buffers that each appeared once, and 8 recency buffers that each appeared once. Therefore, the total number of positions in a given list/block was 356. [(4(20*3)) + 80 + 20 + 8 + 8 = 356].

such that the average S2 position for targets of each condition was statistically equivalent. Third, the corresponding S1 positions and final test positions for all targets were filled. Finally, any positions that were still empty were filled with studied non-targets or tested non-targets. The MATLAB script successfully created four stimulus order outputs, one for each list/block of the experiment.

Using the four order outputs, the experiment was programmed in E-Prime. List/block order was counterbalanced across participants using a Latin square design. Background colour order (between S1/S2) was counterbalanced across trials. Of the 20 target words assigned to the massed, consistent study condition in a given list, 10 appeared on the orange background at S1 and S2 and 10 appeared on the green background at S1 and S2. Of the 20 target words assigned to the massed, variable study condition in a given list, 10 appeared on the orange background at S1 and the green background at S2 and 10 appeared on the green background at S1 and the orange background at S2. Of the 20 target words assigned to the distributed, consistent study condition in a given list, 10 appeared on the orange background at S1 and S2 and 10 appeared on the green background at S1 and S2. Of the 20 target words assigned to the distributed, variable study condition in a given list, 10 appeared on the orange background at S1 and the green background at S2 and 10 appeared on the green background at S1 and the orange background at S2. Of the 20 non-target words in a given list, 10 appeared on the orange background and 10 appeared on the green background. Of the 16 buffers in a given list, 8 appeared on the orange background and 8 appeared on the green background.

Procedure

After consenting to participate in the experiment and completing the colour vision test, participants were capped for the EEG recording. They were seated approximately 50 cm from an 18” CRT computer monitor in a dimly lit room[4]. All personal electronics were powered off. To familiarize participants with the experimental paradigm, they were guided through a practice block. The experimenter explained that words would appear one-at-a-time on the monitor on either an orange background or a green background and that sometimes words might be repeated. These “coloured” trials were described as “study” trials and participants were told they did not need to make a response but simply study the words. The experimenter further explained that sometimes participants would be cued (by way of a fixation cross) to a red word presented on a black background. These “black” trials were described as “test” trials and participants were told to make a recognition response (old/new) using a keypad. Participants completed the practice block of the task with coaching from the experimenter to ensure they understood the task instructions and to get a sense of the timing of stimulus presentation. Pilot testing confirmed that a participant’s hands should remain ready to make a response; therefore, a wrist pad was installed in front of the keypad as a comfort measure. Participants took a break after the practice block and the experimenter ensured that all equipment was recording properly. The first block of the experiment began. Each block lasted approximately 15 minutes and participants were given a 5-minute break in between blocks to rest their eyes.

In each block of the experiment, study trials (i.e., S1 and S2) were characterized as words appearing one-at-a-time at the centre of the screen, in black 18-pt. Arial caps

---

[4] In ERP research, CRT (cathode ray tube) monitors are preferred over LCD (liquid crystal display) monitors because they eliminate image tearing.

font, for a duration of 1500 ms. Words appeared on either the orange background or the green background. A jittered interstudy interval of 750-1250 ms separated trials. The colour of the background during the interstudy interval was the background colour for the next study trial. Target words repeated either 0 or 8 intervening items after their first presentation on either the same or a different coloured background. Test trials (i.e., final test) occurred 20 items after each target word's S2 presentation. Test trials were cued by a white fixation cross presented on a black background for 500 ms followed by a jittered interstudy interval (black screen) of 250-750 ms. The test word then appeared at the centre of the screen, in red 18-pt. Arial caps font, on a black background, for a duration of 1500 ms. Participants were instructed to respond old/new to test trials by pressing one of two labelled keys on the keypad. Handedness of response keys was counterbalanced across participants. In addition to the continuous EEG data, final test data were recorded for behavioural accuracy and RTs. Figure 3 provides an illustration of the paradigm.

Electrophysiological Recording

EEG was recorded for each participant using the BioSemi Active-Two system. Participants were fitted with a mesh cap containing 64 pin-type active Ag-AgCl electrodes positioned according to the International 10-20 system (Figure 4). In addition, two flat-type active electrodes were applied to the participant's left and right mastoid bones. The major advantage of using active electrodes in ERP research (as opposed to passive electrodes) is that they amplify the EEG signal directly at the source, causing an increase in the signal-to-noise ratio. In this way, contamination of the EEG signal from other sources of electricity in the testing room is sufficiently minimized. Scalp impendence for each electrode was kept below 40 kΩ or otherwise noted in the session

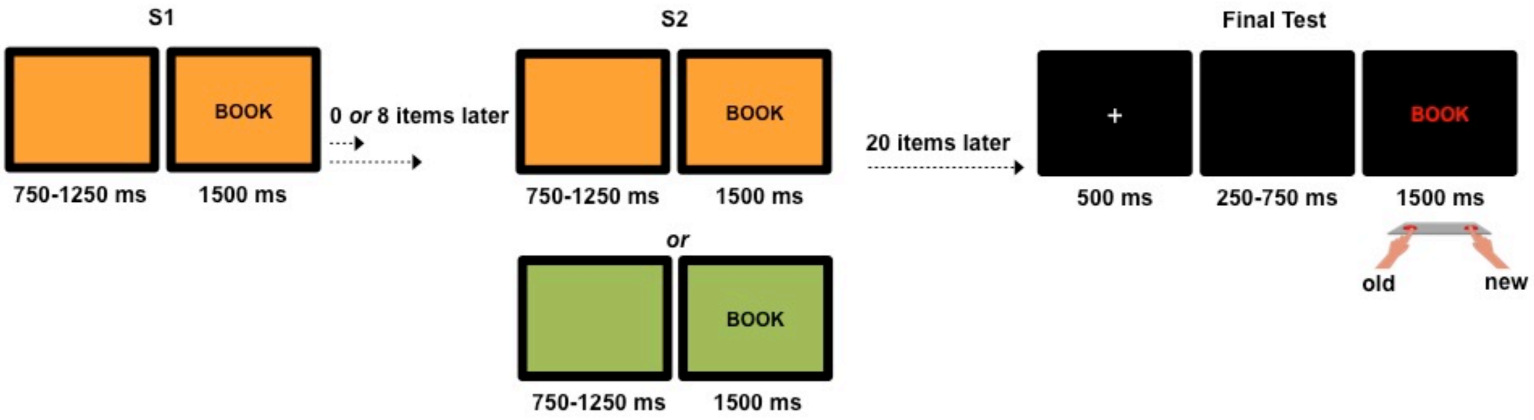

Figure 3. The cued recognition paradigm used in Experiment 1. Words were presented for study (S1) and repeated for study (S2) after a lag of 0 (massed) or 8 (distributed) intervening items on either the same coloured background (consistent study context) or a different coloured background (variable study context) as their first presentations. After a retention interval of 20 intervening items, word recognition was tested. ERPs were time-locked to the onset of the word stimulus.

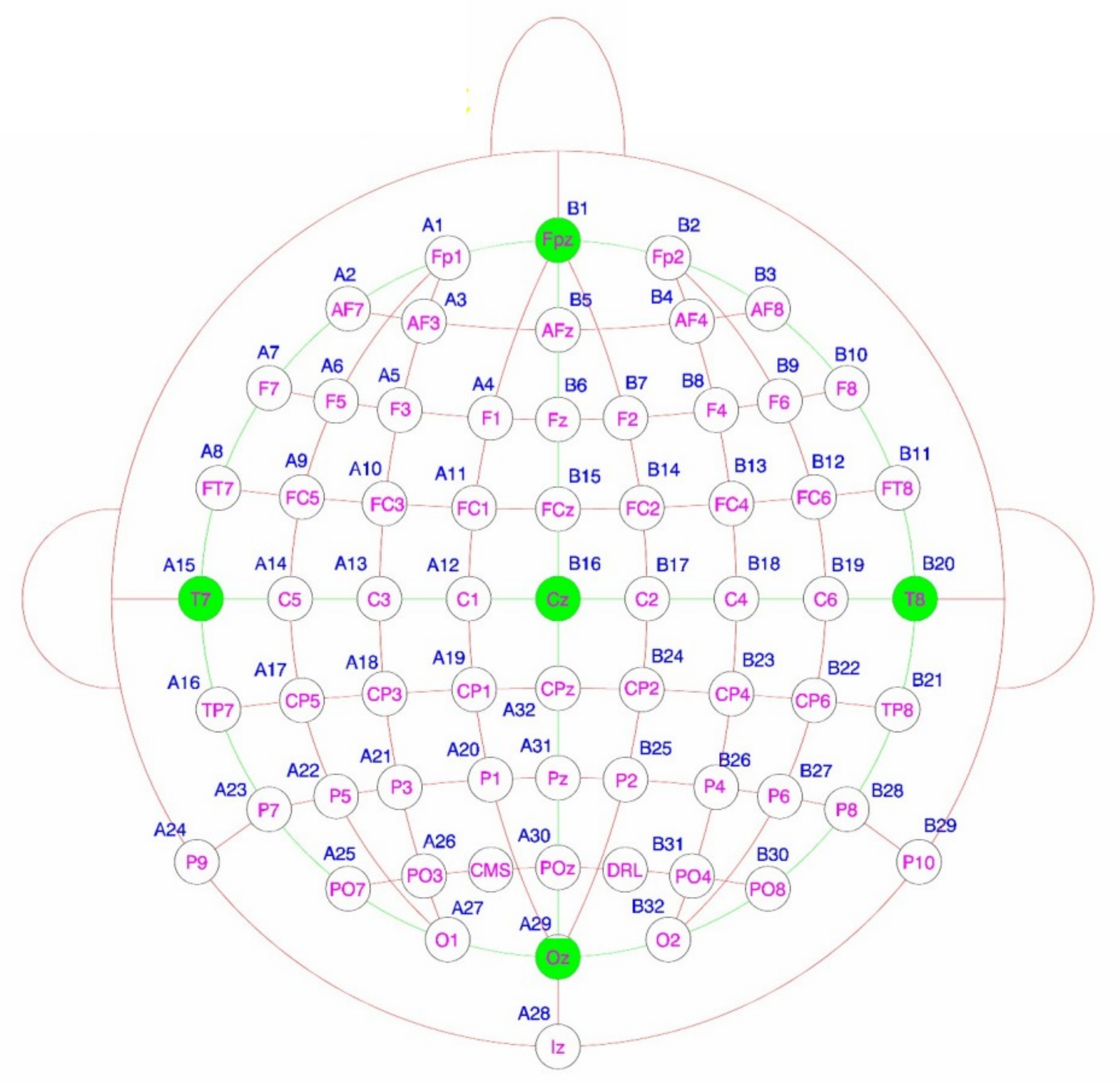

Figure 4. Electrode positions for the BioSemi Active-Two 64-channel system. Reproduced from https://www.biosemi.com/headcap.htm. Two exogenous electrodes were added for the left mastoid and right mastoid.

log for offline pre-processing strategies. During recording, the EEG was sampled at a rate of 2048 Hz, referenced to the Common Mode Sense (CMS) electrode, and amplified using a 0.1-100 Hz bandpass filter. After recording, the data were pre-processed and analyzed using EEGLAB and the ERPLAB plug-in. These programs run in MATLAB.

During pre-processing, the EEG data were down-sampled to a rate of 512 Hz, re-referenced to the linked mastoids, and filtered using a bandpass of 1-30 Hz. This frequency range is known to capture the ERP components of interest in most cognitive neuroscience experiments while successfully filtering out high frequency artifacts (e.g., muscle contractions) and low frequency artifacts (e.g., sweating). By way of visual inspection, periods of blank time and/or unfiltered muscle artifacts were flagged and removed. Noisy channels were also flagged. Specific artifacts associated with eye movements (blinks and saccades) and heartbeat (pulse) were detected using independent components analysis (ICA) from all clean channels. On a participant-by-participant basis, independent components that were characterized as large spikes and/or square waves of electrical activity in frontal regions of the scalp (eye movements) or small, regularly occurring waves of electrical activity in mastoid/lateral temporal regions of the scalp (heartbeat) were removed from the continuous EEG. After ICA, any channels previously identified as noisy were interpolated instead. The continuous EEG was segmented into 1000 ms epochs and baseline-corrected relative to a 200 ms pre-stimulus period. The segmented data were manually inspected again to remove any outstanding contaminated trials. Average ERPs were computed for each participant for the following conditions: S1 targets (per condition, all trials); S2 targets (per condition, all trials); final test targets (per condition, correct trials only); S1 studied, non-targets (all trials); and S1 tested, non-targets (correct trials only). Table 2 summarizes the average number of artifact-free trials per condition that were included in the various ERP analyses. On the basis of the literature previously summarized, and visual inspection of the data, mean amplitude and/or latency analyses were conducted for specific electrodes during specific time windows. Specifically, the visual N1 was assessed at P5 between 150-225 ms, the FN400 at Fz between 275-375 ms (for study trials) and 300-500 ms (for test trials), and the LPC at Pz between 400-1000 ms. Unless otherwise stated, the specific analysis conducted was a 2 × 2 repeated measures ANOVA. Bayes factors were also calculated to confirm main effects and interactions and clarify any marginally significant results.[5]

Results

Behavioural Data (Final Test)

Behavioural data includes participants' corrected recognition accuracy to test trials (hits minus false alarms, reported as percentage correct) and RTs to test trials (correct trials only, reported in ms).

Distributed practice effects. For the accuracy analysis (Figure 5a), there was a significant main effect of Lag, $F(1, 15) = 26.44$, $p < .001$, $\eta_p^2 = .64$, $\mathrm{BF}_{\mathrm{Inclusion}} = 329.80$ (extreme evidence for H_1). Words studied at the distributed lag ($M = 62.79$, $SD = 15.40$)

---

[5] Null hypothesis significance testing (NHST; e.g., ANOVA) tests only whether to accept or reject a null hypothesis (H_0); it implies nothing about an alternative hypothesis (H_1). On the contrary, Bayesian statistics test the probability of the observed data under H_1 relative to H_0, providing a richer interpretation of the data. For a simple interpretation, a BF value of 1 means the data are equally probable under H_0 and H_1. A BF value less than 1 means the data are more probable under H_0 relative to H_1 (0.33-1 = anecdotal evidence for H_0; 0.1-0.33 = moderate evidence for H_0; <0.1 = strong evidence for H_0). A BF value greater than 1 means the data are more probable under H_1 relative to the H_0 (1-3 = anecdotal evidence for H_1; 3-10 = moderate evidence for H_1; >10 strong evidence for H_1). For the interested reader, see Jarosz and Wiley (2014).

Table 2

Number of Artifact-Free ERP Trials By Condition

Condition	S1		S2		Final Test	
	M	*SD*	*M*	*SD*	*M*	*SD*
Massed, consistent study	64.88	7.06	63.25	9.14	43.44	7.30
Massed, variable study	64.00	8.63	65.25	7.78	46.75	10.47
Distributed, consistent study	63.56	7.89	65.56	7.23	53.56	10.87
Distributed, variable study	62.81	7.94	65.69	7.60	50.75	9.47
Non-targets, studied	62.63	9.14	--	--	--	--
Non-targets, tested	--	--	--	--	60.25	13.57

Note. For S1 and S2, all trials were analyzed. For final test, correct trials were analyzed.

were more accurately recognized after the retention interval than words studied at the massed lag (*M* = 56.50, *SD* = 16.06). The main effect of Study Context and the Lag × Study Context interaction were not significant.

For the RT analysis (Figure 5b), there was a significant main effect of Lag, $F(1, 15) = 29.59$, $p < .001$, $\eta_p^2 = .66$, $\mathrm{BF}_{\mathrm{Inclusion}} = 3925.09$ (extreme evidence for H_1). Participants were faster to respond to test words that had been studied at the distributed lag (*M* = 865, *SD* = 114) compared to test words that had been studied at the massed lag (*M* = 905, *SD* = 118). The main effect of Study Context was not significant. There was a significant Lag × Study Context interaction, $F(1, 15) = 4.79$, $p < .05$, $\eta_p^2 = .24$, $\mathrm{BF}_{\mathrm{Inclusion}} = 1.26$ (anecdotal evidence for H_1). Post-hoc paired samples *t*-tests showed that the RT difference between massed and distributed targets was more robust for targets that had been consistently studied [$t(15) = 4.70$, $p < .001$, $d = 1.20$, $\mathrm{BF}_{10} = 114.23$ (extreme evidence for H_1)] compared to variably studied [$t(15) = 3.06$, $p < .01$, $d = 0.85$, $\mathrm{BF}_{10} = 6.65$ (moderate evidence for H_1)].

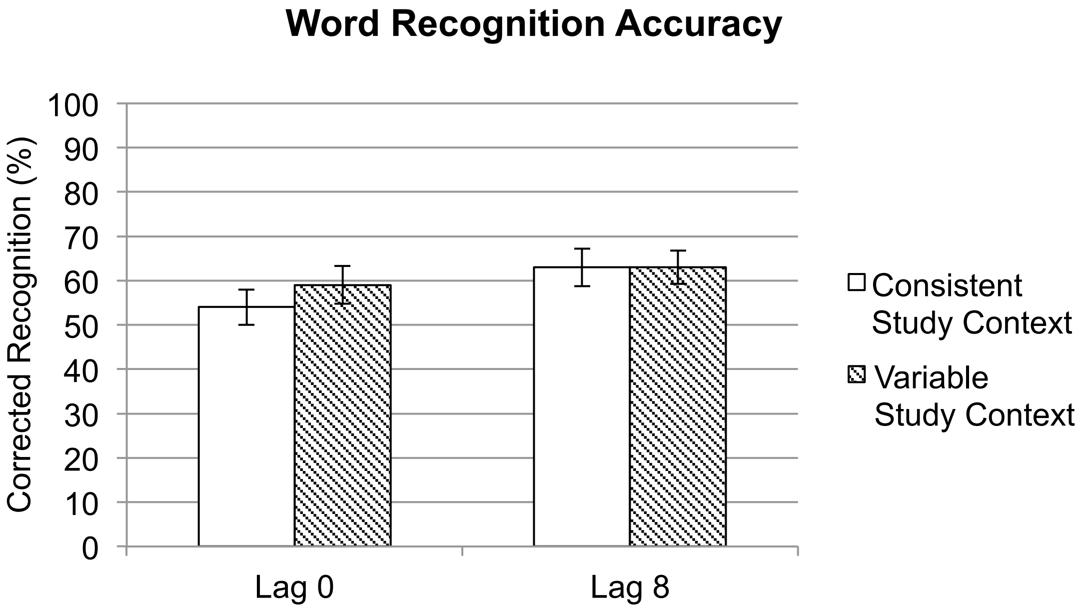

Figure 5a. Participants' corrected accuracy scores (hits minus false alarms) in the cued recognition task. Error bars represent standard error of the mean.

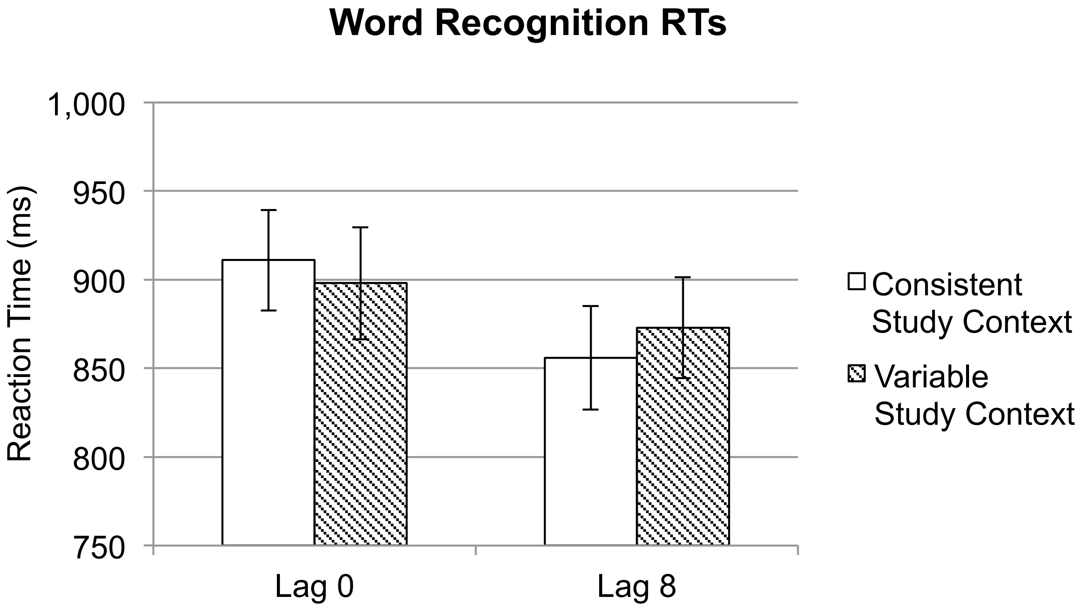

Figure 5b. Participants' reaction time to correctly recognized targets in the cued recognition task. Error bars represent standard error of the mean.

ERP Data (S1, S2, and Final Test)

ERP data includes participants' mean amplitude measurements (reported in µV) and/or peak latency measurements (reported in ms) at a specific electrode and time window of interest. For S1 and S2, all trials were analyzed. For final test, correct trials were analyzed.

Old/new repetition effect. To confirm replication of the ERP old/new repetition effect, ERPs to first and second presentations of words were compared by way of visual inspection. Target words repeated at S2 (collapsed across condition) elicited more positive-going ERPs approximately 300 ms post-stimulus compared to their first presentations at S1 (collapsed across condition) and to non-targets studied for the first/only time. The effect was observed across the entire scalp and especially at central sites (Figure 6).

Visual N1 at S2. Mean amplitude of the visual N1 component at S2 was examined at electrode P5 between 150-225 ms. The main effect of Lag was not significant. The main effect of Study Context was significant, $F(1, 15) = 7.21, p < .05$, $\eta_p^2 = .33$, $BF_{Inclusion} = 2.94$ (moderate evidence for H_1), demonstrating greater visual N1 attenuation for consistently studied targets at S2 ($M = 0.42$, $SD = 1.86$) compared to variably studied targets ($M = -0.40$, $SD = 1.74$; Figure 7a). The Lag × Study Context interaction was not significant.

FN400 at S2. Mean amplitude of the FN400 component at S2 was examined at electrode Fz between 275-375 ms. There was a marginally-significant effect of Lag, $F(1, 15) = 3.78, p = .07$, $\eta_p^2 = .20$, $BF_{Inclusion} = 0.24$ (moderate evidence for H_0), suggesting a trend of greater FN400 mean amplitude (i.e., less attenuation) for distributed targets at S2

(*M* = -2.09, *SD* = 2.90) compared to massed targets (*M* = -1.12, *SD* = 2.74; Figure 7b). Since the Bayes factor opposed the Lag trend, it was not considered further. The main effect of Study Context was not significant according to NHST (*p* = .09); however, it had a moderate effect according to Bayes ($BF_{Inclusion}$ = 3.03). FN400 mean amplitude was attenuated for consistently studied targets at S2 (*M* = -1.25, *SD* = 2.20) relative to variably studied targets (*M* = -1.95, *SD* = 2.81). The Lag × Study Context interaction was not significant.

LPC at S2. Mean amplitude of the LPC component at S2 was examined at electrode Pz between 400-1000 ms. There was a marginally-significant effect of Lag, $F(1, 15) = 3.95$, $p = .07$, $\eta_p^2 = .21$, $BF_{Inclusion} = 1.40$ (anecdotal evidence for H_1), representing a trend of greater LPC mean amplitude for massed targets at S2 (*M* = 1.33, *SD* = 2.14) compared to distributed targets (*M* = 0.60, *SD* = 2.05). There was a significant main effect of Study Context, $F(1, 15) = 4.70$, $p < .05$, $\eta_p^2 = .24$, $BF_{Inclusion} = 1.73$ (anecdotal evidence for H_1). LPC mean amplitude was greater for consistently studied targets at S2 (*M* = 1.35, *SD* = 2.35) compared to variably studied targets (*M* = 0.58, *SD* = 1.79). The Lag × Study Context interaction was not significant.

Peak latency of the LPC component at S2 was also examined at electrode Pz between 400-1000 ms. There was a significant main effect of Lag, $F(1, 15) = 5.55$, $p < .05$, $\eta_p^2 = .27$, $BF_{Inclusion} = 2.31$ (anecdotal evidence for H_1). The LPC took longer to peak for distributed targets at S2 (*M* = 764, *SD* = 151) compared to massed targets (*M* = 671, *SD* = 178). The main effect of Study Context and the Lag × Study Context interaction were not significant (Figure 7c).

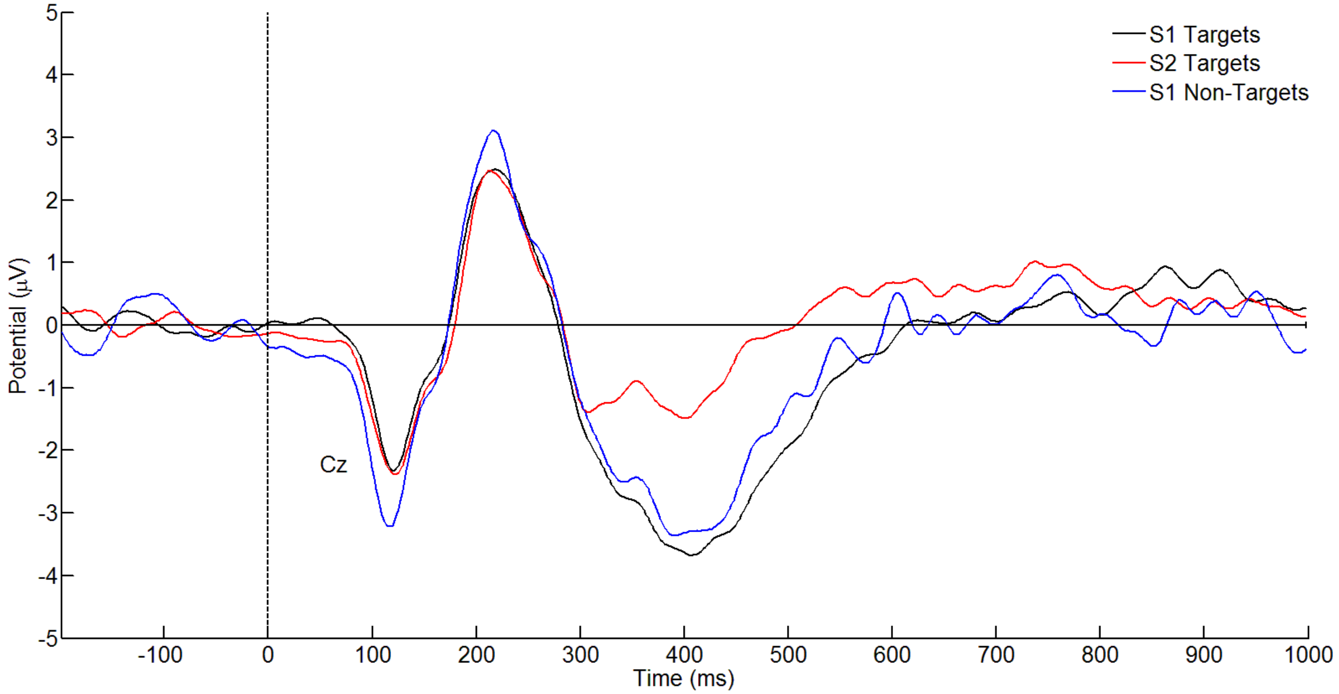

Figure 6. An ERP old/new repetition effect. ERPs to target words at S2 were more positive-going ~300 ms post-stimulus compared to ERPs to target words at S1 (and to non-targets studied for the first/only time).

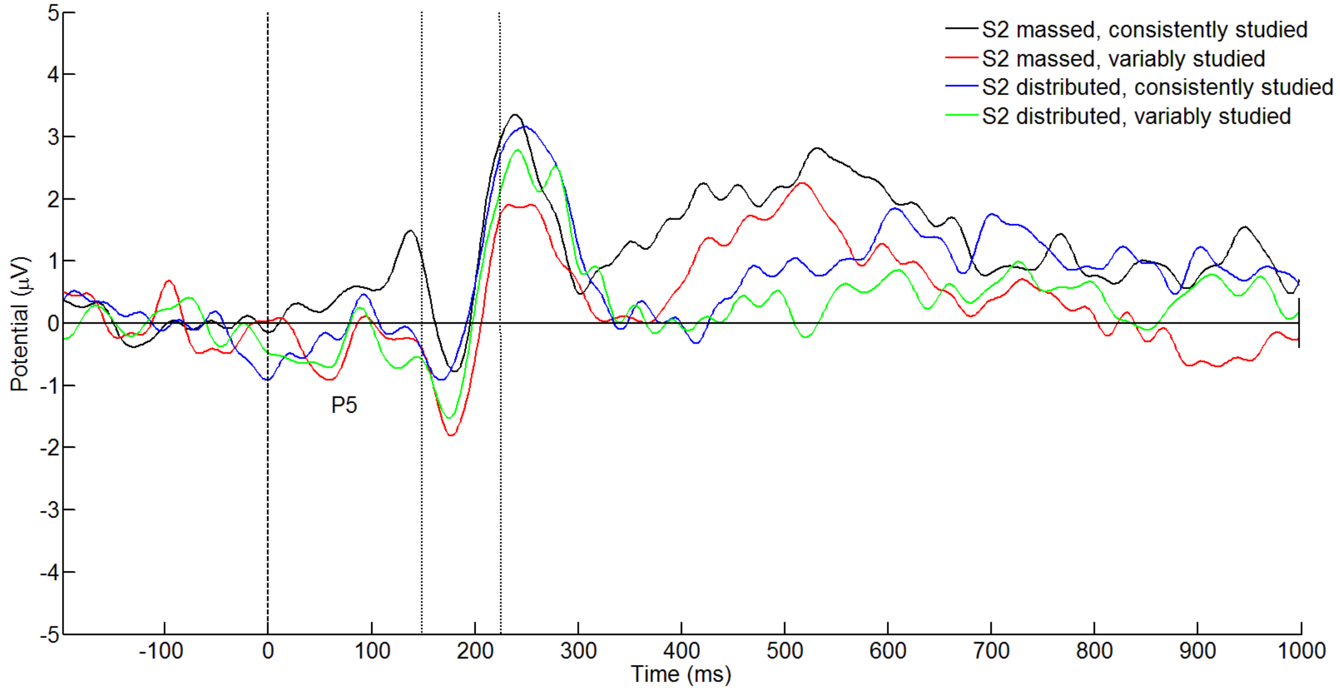

Figure 7a. The visual N1 component at S2. Consistently studied targets elicited an attenuated visual N1 compared to variably studied targets ($p < .05$; $BF_{Inclusion} = 2.94$, moderate evidence for H_1).

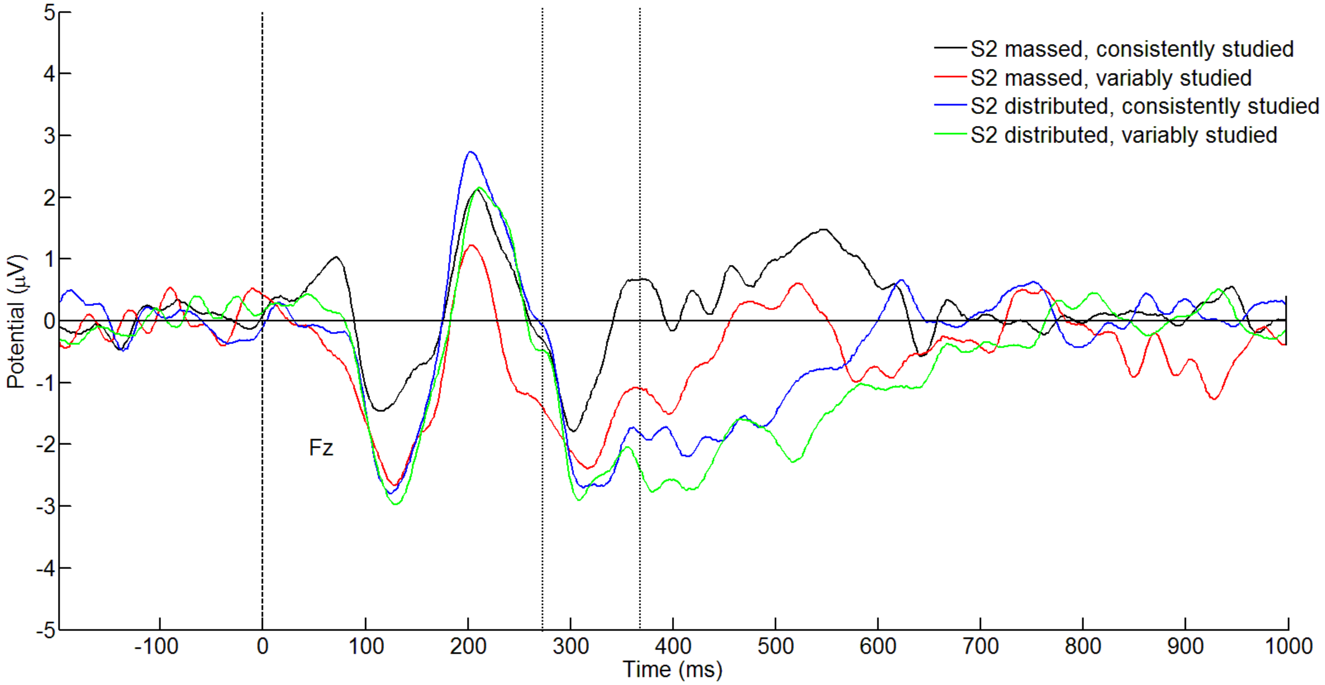

Figure 7b. The FN400 component at S2. There were trends of greater FN400 mean amplitude (i.e., less attenuation) for distributed targets compared to massed targets (p = .07; $BF_{Inclusion}$ = 0.24, moderate evidence for H_0) and for variably studied targets compared to consistently studied targets (p = .09; $BF_{Inclusion}$ = 3.03, moderate evidence for H_1).

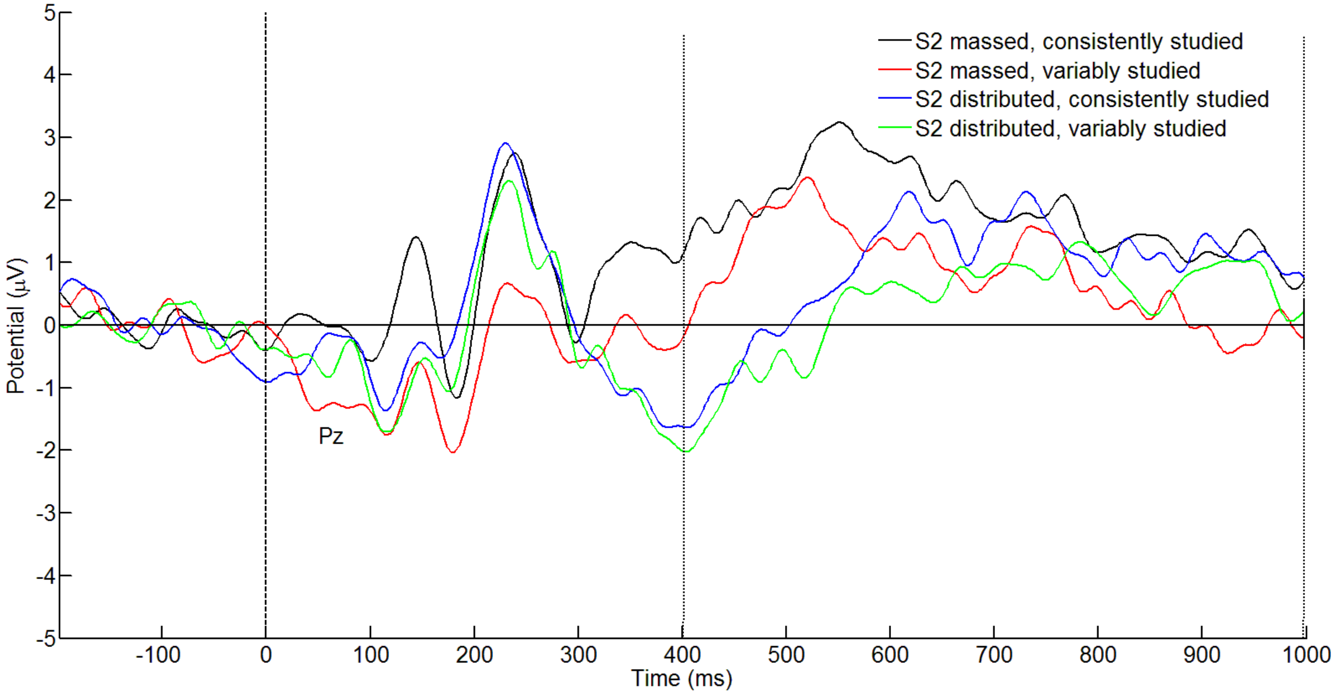

Figure 7c. The LPC component at S2. There was a trend of greater LPC mean amplitude for massed targets compared to distributed targets ($p = .07$; $BF_{\text{Inclusion}} = 1.40$, anecdotal evidence for H_1). The component was also larger in mean amplitude for consistently studied targets compared to variably studied targets ($p < .05$; $BF_{\text{Inclusion}} = 1.73$, anecdotal evidence for H_1) and took longer to peak for distributed targets compared to massed targets ($p < .05$; $BF_{\text{Inclusion}} = 2.31$, anecdotal evidence for H_1).

FN400 and LPC at final test. Mean amplitude of the FN400 at final test was examined at electrode Fz between 300-500 ms. Targets were first compared to non-targets using a series of paired samples *t*-tests. Mean amplitude of the FN400 was attenuated for distributed, variably studied targets ($M = -0.44$, $SD = 4.34$) compared to non-targets ($M = -2.77$, $SD = 4.26$), $t(15) = 2.43$, $p < .05$, $d = 0.61$, $BF_{10} = 2.37$ (anecdotal evidence for H_1); however, there were no significant differences in FN400 mean amplitude between non-targets and the other target conditions (Figure 8a).

Mean amplitude of the FN400 at final test was also compared exclusively amongst the four target conditions. This analysis revealed no significant main effects; though, there was a marginally-significant Lag × Study Context interaction, $F(1, 15) = 10.48$, $p = .07$, $\eta_p^2 = .19$, $BF_{\text{Inclusion}} = 0.37$ (anecdotal evidence for H_0). This trend suggested that varying a target's study context at S2 attenuated the FN400 for distributed targets at final test ($MD = +1.31$) but not massed targets ($MD = -0.31$). Since the Bayes factor opposed the interaction trend, it was not considered further.

Mean amplitude/peak latency of the LPC at final test was examined at electrode Pz between 400-1000 ms. Targets were first compared to non-targets using a series of paired samples *t*-tests. There were no significant differences in LPC mean amplitude or peak latency between non-targets and any target condition (Figure 8b).

Mean amplitude/peak latency of the LPC at final test was also compared exclusively amongst the four target conditions. The mean amplitude analysis revealed a significant main effect of Lag, $F(1, 15) = 18.54$, $p < .01$, $\eta_p^2 = .55$, $BF_{\text{Inclusion}} = 12.62$ (strong evidence for H_1). LPC mean amplitude was larger for distributed targets at final test ($M = 4.03$, $SD = 2.61$) compared to massed targets ($M = 3.36$, $SD = 2.66$). The main

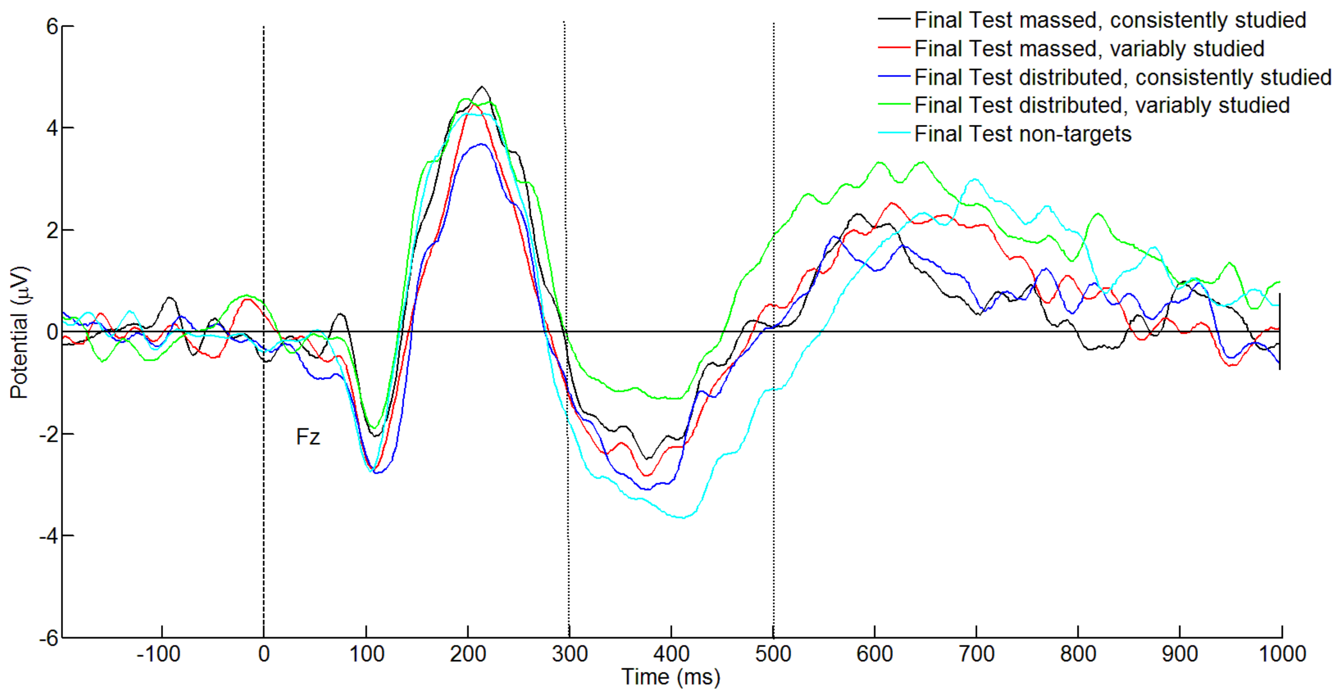

Figure 8a. The FN400 component at final test. Compared to correctly recognized non-targets, mean amplitude of the FN400 was attenuated only for correctly recognized distributed, variably studied targets ($p < .05$; $BF_{10} = 2.37$, anecdotal evidence for H_1).

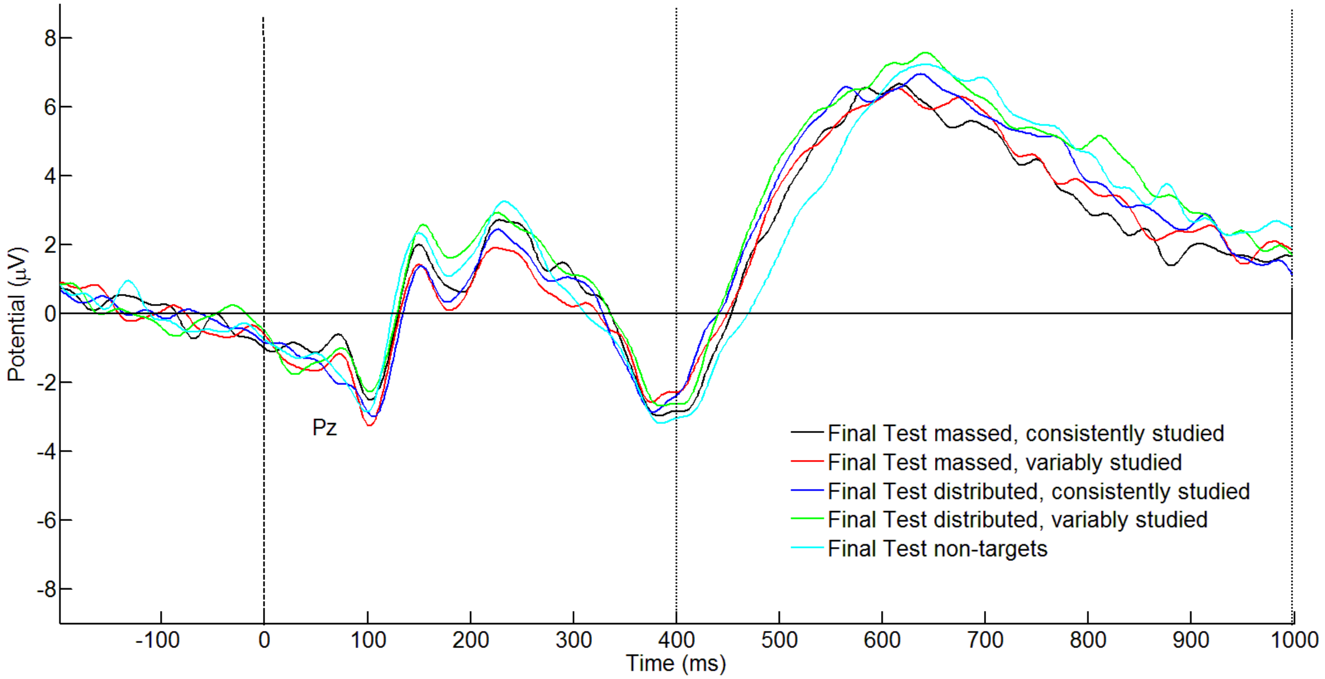

Figure 8b. The LPC component at final test. There were no significant differences in LPC mean amplitude/peak latency between non-targets and any target condition.

effect of Study Context and the Lag × Study Context interaction were not significant. The peak latency analysis had no significant findings.

Discussion

The main purpose of this experiment was to investigate the neural basis of the distributed practice effect using continuous EEG as well as whether exogenous factors—in this case, study context—influence this neural representation. In a traditional paradigm, words were presented for study only once or they were presented twice, in either a massed or distributed fashion, on either the same or a different coloured background. Several intervening items later, recognition of the words was tested.

According to the hybrid account, deficient processing, encoding variability, and/or study-phase retrieval mechanisms work in concert to account for the advantage of distributed practice. More specifically, relative to massed repetitions at S2, distributed repetitions at S2 are hypothesized to (a) receive superior reprocessing (b) be associated with a greater number of unique contextual cues and/or (c) elicit greater retrieval effort. Any or all of these contributing factors may explain why, in a final test scenario, repetitions studied using a distributed schedule are typically better remembered than repetitions studied using a massed schedule. Although the distributed practice effect is a robust memory phenomenon, a number of behavioural studies demonstrate a weaker advantage of distributed practice when to-be-learned items are repeated in variable study contexts (i.e., a Lag × Study Context interaction; Table 1). The hybrid account concurs that massed items should no longer suffer from deficient processing under variable study conditions and, although distributed items should benefit from an even greater degree of encoding variability under variable study conditions (i.e., the drift of time *and* a change in study environment), their study-phase retrieval may be less successful. Therefore, distributed practice is theorized to be most effective under consistent study conditions, a finding that should be evident both behaviourally (i.e., final test performance) and neurally (i.e., ERPs during study/test).

Behavioural Evidence of the Distributed Practice Effect

The behavioural data of the present study can be summarized by three findings. First, there was a distributed practice effect. Words studied at the distributed lag were subsequently recognized faster and with greater accuracy on test trials relative to words studied at the massed lag. Second, irrespective of lag, there was no particular advantage to studying a target in a consistent study context or variable study context. Finally, there was mixed support for the Lag × Study Context interaction. On the one hand, study context did not significantly interact with lag to influence participants' final recognition accuracy. As illustrated in Figure 5a, recognition of massed targets was slightly improved if they had been studied on two different coloured backgrounds; however, ultimately, recognition of distributed targets (consistently or variably studied) was superior. On the other hand, study context did interact with lag to influence participants' recognition speed. As illustrated in Figure 5b, RTs to massed targets were faster if they had been studied in a variable study context relative to a consistent study context and RTs to distributed targets were slower if they had been studied in a variable study context relative to a consistent study context. This outcome was realized statistically as a smaller RT difference between massed, variably studied targets and distributed, variably studied targets compared to the RT difference observed between massed, consistently studied targets and distributed, consistently studied targets. Although the Lag × Study Context interaction was apparent in the behavioural RT data, it was not a robust finding. Hypothesis 1 was partially supported.

It is possible that the interaction effect was harder to detect in the present experiment because it was a recognition paradigm. Indeed, the Lag × Study Context interaction is more frequently reported in free recall paradigms. According to Glenberg (1979), in a free recall scenario where there are no salient test cues, a stimulus characterized by many unique cues from the study phase—including descriptive, structural, and/or contextual cues—is most likely to be recalled due to a greater probability that at least one or more of these study cues overlaps with randomly fluctuating cues in the test environment. Variation in study context between S1 and S2 should be helpful for massed targets and (possibly) distributed targets. Moreover, distributed targets should preferentially benefit from a greater number of unique temporal cues. In a recognition scenario, however, the outcome may be different because the test cue is extremely salient. Successful item recognition requires only that a single descriptive study cue match the test cue (i.e., the word itself). Additional cues from the study phase—in this case, study context effects and lag effects—become less influential to the retrieval process under recognition demands. Although Glenberg's rationale has received some empirical support over the years, a number of studies do report a Lag × Study Context interaction using recognition tasks (e.g., Hintzman, Summers, & Block, 1975; Rose, 1980). A recognition task was chosen for the present experiment so that ERPs could be time-locked to words presented at S1, S2, and final test. In summary, the behavioural data support distributed repetition of verbal stimuli over massed repetition—in any study context—as a means of improving subsequent recognition accuracy. Study context may only influence the magnitude of the distributed practice advantage in recognition when behaviour is assessed using a more sensitive measure, such as reaction time.

Neural Evidence of the Distributed Practice Effect

The ERP data of the present study can be summarized by a number of interesting findings. A manipulation check confirmed that there was a clear ERP old/new repetition effect, which supported Hypothesis 2. ERPs to second presentations of targets were more positive-going than ERPs to first presentations of the same targets and to non-targets presented only once. This finding exemplifies how ERPs can suggest cognitive processing differences between two conditions in the absence of an overt behavioural response. It also confirms that participants were engaged in the experimental task.

To measure differences in neural responses to words repeated at different lags, and in different study contexts, the visual N1, FN400, and LPC components elicited at S2 were compared across conditions. First, mean amplitude of the visual N1 at S2 was attenuated for consistently studied targets, which may represent an electrophysiological form of perceptual priming (e.g., Wiggs & Martin, 1998). Mollison (2015) reported a visual N1 lag effect where massed targets in his paradigm evoked greater visual N1 attenuation compared to distributed targets. This finding was not replicated in the present data, nor did massed, consistently studied targets exhibit the greatest visual N1 attenuation compared to all other targets. Therefore, deficient processing mechanisms of the distributed practice effect did not have a clear neural representation in the present experiment, at least not indexed by the visual N1. Although this outcome does not provide support for Hypothesis 3, it should be noted that there were interesting waveform differences among target conditions at S2 that preceded the visual N1, most clearly, the P1 component. The P1 component indexes bottom-up sensory processing of a stimulus within the first 100 ms of its onset. More specific to the present paradigm, the P1 captures neuronal activity in the extrastriate cortex, which is sensitive to colour (Pratt, 2012). Closer examination of Figure 7a shows that massed, consistently studied targets—which were words repeated back-to-back on the same coloured background—elicited a larger and later-to-peak P1 compared to the other target conditions. This outcome might reflect a sort of “stimulus persistence” EEG signal, which was unique to massed, consistently studied targets. Although this is one interpretation of the P1 data, it does run counter to standard explanations of deficient processing in terms of repetition priming/suppression (where stimulus persistence is associated with *decreases* in neural signal; e.g., Schacter, Wig, & Stevens, 2007).

Second, mean amplitude of the FN400 at S2 was attenuated for consistently studied targets, which may indicate that background colour triggered the familiarity of repeated targets regardless of their lag. The ERP data did not clearly replicate previous research documenting greater FN400 mean amplitude for distributed repetitions at S2 compared to massed repetitions. The FN400 lag effect has been described as a neural representation of encoding variability mechanisms of distributed practice, where distributed targets are experienced as less familiar at S2 compared to massed targets, therefore eliciting little/no FN400 attenuation (Kim et al., 2001; Mollison, 2015; Van Strien et al., 2007). Although the data trended in the predicted direction, evidence in support of the FN400 lag effect in the present experiment was indeterminate according to

Bayesian statistics. The FN400 Lag × Study Context interaction was also not significant. Therefore, Hypothesis 4 was not supported.

Third, there was a trend of greater LPC mean amplitude for massed targets at S2 compared to distributed targets. This trend was not indeterminate according to Bayesian statistics; however, it was not robust either. The LPC was also larger in mean amplitude for consistently studied targets compared to variably studied targets. One interpretation of these findings is that a larger LPC represents active, sustained maintenance of recently presented items in working memory (e.g., Mollison, 2015). In terms of LPC latency, the LPC peaked significantly later for distributed targets compared to massed targets. This finding replicates previous research (e.g., Kim et al., 2001) and supports a neural representation of study-phase retrieval mechanisms of the distributed practice effect, where distributed items are considered to be more effortful to retrieve at S2 compared to massed items (and where retrieval effort is indexed by a later-to-peak ERP component). Once again, however, this finding was not particularly robust. The fact that the LPC latency lag effect did not interact with study context suggests that variable study did not further weaken retrievability of distributed items. In other words, study-phase retrieval of distributed, consistently studied targets and distributed, variably studied targets were similarly effortful (and more effortful than all massed items). Overall, Hypothesis 5 was partially supported.

Finally, the FN400 and LPC components were compared across conditions again at the final test. Contrary to expectations, *t*-test analyses demonstrated that only distributed, variably studied targets elicited a familiarity response (i.e., a difference in FN400 mean amplitude relative to non-targets) and none of the targets elicited a recollection response (i.e., a difference in LPC mean amplitude and/or peak latency relative to non-targets). Characteristics of the experimental paradigm may account for these data. For example, the to-be-learned stimuli in the experiment included hundreds of words, presented quickly, and tested by way of recognition. It is likely that participants engaged in more of a "gist" retrieval strategy when faced with a test trial rather than retrieval with episodic details. In this way, the ERP data complement conclusions made from the behavioural data, namely, that distributed, variably studied targets were associated with superior subsequent (familiarity) recognition. This result provides some support for Hypothesis 6; however, it should be interpreted with caution. Although distributed, variably studied targets elicited FN400 attenuation compared to non-targets, additional analyses concluded that they did not elicit greater FN400 attenuation compared to all other targets. Furthermore, although distributed targets were associated with greater LPC mean amplitude at final test compared to massed targets, they were not associated with LPC mean amplitude differences in comparison to non-targets. These two additional findings may call into question the robustness of the neural recognition response elicited by distributed targets on the final test.

To summarize, the ERP data replicated some of the findings from previous studies reporting a neural distinction between massed and distributed repetitions at S2, although, for the majority of the analyses, the effects were weak. Differences in the LPC waveform measured at S2 suggested that massed repetitions were less likely to be semantically re-encoded at S2 because they were still in working memory and that distributed repetitions may have engaged greater retrieval effort at S2. These results represent electrophysiological evidence that is convergent with study-phase retrieval explanations of the distributed practice effect (e.g., Thios & D'Agostino, 1976). Neural evidence of deficient processing and encoding variability mechanisms were less clear from the S2 ERP data, which could be the result of insufficient experimental power. In the absence of specific prior evidence, the ERP power analysis borrowed effect sizes from behavioural studies to estimate the sample size needed to detect electrophysiological effects with 95% power. However, if neural distributed practice effects are in fact smaller than behavioural distributed practice effects, as the present results suggest, additional data were needed. Based on a revised power analysis using a medium effect size, it is recommended that future studies investigating an ERP Lag × Study Context interaction maintain at least 20 participants.

The ERP data also extended previous research by examining whether a neural distinction between massed and distributed repetitions remained when items were correctly recognized after a retention interval. Differences in the FN400 waveform at final test suggested that distributed, variably studied targets were recognized as being more familiar than non-targets (though not necessarily more familiar than other correctly recognized targets). Although there were no specific differences in the LPC waveform at final test between any target condition and non-targets, the nature of the recognition paradigm may account for this finding.

Finally, the ERP data weighed in on the question of whether study context influences the neural basis of the distributed practice effect. A weak Lag × Study Context interaction was apparent only for participants' RTs to test trials. An advantage of ERP methodology is that it is sensitive to the timing of cognitive processing; therefore, complementing the behavioural RT data, a Lag × Study Context interaction was hypothesized across all ERP components of interest. Instead, there were no significant interaction effects. This general finding is not surprising given the lack of an obvious interaction in the behavioural data. In addition to the possibility that a recognition paradigm weakened the interaction effect, it is also possible that the study context manipulation was not robust/relevant enough to elicit the interaction effect. A variably studied item may have appeared on a different coloured background at S2 compared to S1; however, this "variable" background colour was not different from other repeated targets (i.e., there were only two background colours). Thus, the effect of study context on the neural correlates of the distributed practice effect is inconclusive using the current paradigm. For stronger conclusions to be drawn, it is recommended that future ERP investigations retest the interaction using a free recall paradigm and/or a different type of study context manipulation (e.g., stimulus-response pairs, using consistent/variable stimulus words).

Study 2

Despite empirical evidence across a variety of different paradigms, research investigating the effect of study context on the distributed practice effect remains limited in developmental scope. The literature is biased towards university samples (e.g., all of the studies reported in Table 1) with the exception of one study confirming a Lag × Study Context interaction in elementary school children (Toppino & DeMesquita, 1984). The interaction effect has not been explored in older adults. An outstanding question is whether certain age-related memory impairments, such as older adults' difficulty associating items with their respective study contexts (e.g., Old & Naveh-Benjamin, 2008), affect how study context influences the distributed practice effect in an aging population. In Study 2, older adult participants aged 60+ years were compared to younger adult participants in terms of their performance on a distributed practice free recall paradigm. Items repeated at a massed lag or one of two distributed lags on either the same or a different background scene. This study is among the first to test whether age moderates the Lag × Study Context interaction effect.

Distributed Practice Benefits in Aging

Previous research confirms that older adults benefit from traditional distributed practice. For example, in a study by Balota, Duchek, and Paullin (1989), younger and older adults learned unrelated word pairs repeated at lags of 0, 1, 4, 8, or 20 intervening items and tested (cued recall) after a retention interval of 20 intervening items. Although older adults remembered fewer items overall compared to younger adults, distributed practice improved memory performance for both groups (a 9% advantage of spacing in younger adults and a 5% advantage of spacing in older adults, as stated by the authors).

As noted by Balota and colleagues (1989), and replicated by others (e.g., Simone, Bell, & Cepeda, 2012), the magnitude of the distributed practice effect is often smaller for older adults than it is for younger adults. Balota explained this discrepancy from an encoding variability perspective, suggesting that older adults encode fewer contextual elements during learning and/or that contextual elements fluctuate from available to unavailable states more slowly in older age. For these reasons, older adults will not gain the same quantity and variety of contextual cues from distributed repetition as those gained by younger adults. Weaker distributed practice benefits in aging can also be explained from a study-phase retrieval perspective, specifically that older adults experience greater study-phase retrieval failure during learning than younger adults. This theory receives support from a study by Kilic and colleagues (2013) where younger and older adults studied words repeated lags of 1-3, 6-10, or 40-56 intervening items and provided confidence ratings about whether each word had been previously encountered or not. Confident hits (correct response: "definitely old") decreased monotonically as a function of lag for both age groups. More importantly, there was an Age × Lag interaction, demonstrating that the decrease in hits across lags was steeper for older adults than for younger adults[6]. This finding suggests that older adults are less likely to

---

[6] Recognition decisions are affected by response bias, which is the tendency for a participant to classify an item as "old" or "new" based on whether its memory strength exceeds a criterion. Research suggests that criterion placement may vary with age; however, the results are mixed. Some studies have found that age is associated with a more liberal response bias (i.e., older adults are more likely to endorse an item as "old" compared to younger adults; e.g., Huh, Kramer, Gazzaley, & Delis, 2006), while other studies have found that age is associated with a more conservative response bias (i.e., older adults are more likely to endorse an item as "new" compared to younger adults; e.g., Criss, Aue, & Kilic, 2014). In the case of Kilic et al.'s (2013) findings, response bias differences between age groups were controlled by showing the same results for a

successfully retrieve the S1 occurrence of a distributed item at S2, in turn hindering the effect of distributed practice on subsequent memory.

The distributed practice effect in older adults has been explored in between-session paradigms (e.g., Simone et al., 2012) and in paradigms with more than two study opportunities (e.g., Logan & Balota, 2008). A recent study by Bercovitz, Bell, Simone, and Cepeda (2017) aimed to test the Lag × Study Context interaction with an older adult sample; however, because the experimental paradigm confounded study context and test context, the final test data are difficult to interpret. Furthermore, unlike the present research, the paradigm was a between-sessions design.

Memory Impairments in Aging

Although study context influences the distributed practice effect in younger adults, there is reason to suspect that it may be less influential on the distributed practice effect in older adults. Research that contrasts older adults' memory abilities for an *item* versus its *study context* provides a basis for this hypothesis. According to different meta-analyses (Old & Naveh-Benjamin, 2008; Spencer & Raz, 1995) and longitudinal studies (Cansino et al., 2013), age differences in study context memory are reliably greater than those in item memory. A classic study by Schacter and colleagues (1991) demonstrates one example of this relative impairment. Younger and older participants watched a video of two people (a male and a female) speaking aloud fictitious facts about well-known celebrities (e.g., *Bob Hope's father was a fireman*). Participants were instructed to pay attention to both the item (i.e., the fact) and its study context (i.e., the gender of the speaker) as they watched the video. Immediately after the study phase, and again two

---

subsample of younger and older participants that did not differ in terms of false alarms at S1.

hours later, they were tested for their recall memory (*What job did Bob Hope's father have? Did you learn this information from a male or female speaker?*) The results of this study showed that, relative to younger adults, older adults had superior item memory and impaired study context (source) memory at both testing occasions. Related research has reported aging deficits for other contextual features associated with to-be-learned items, such as: background scene (Denney, Miller, Dew, & Levav, 1991), background colour (Park & Puglisi, 1985), spatial location and font (Naveh-Benjamin & Craik, 1995), and case (i.e., uppercase or lowercase; Kausler & Puckett, 1980).

Impaired memory for study context may be related to older adults' difficulty binding information into complex memories during encoding, a theory tested in a series of experiments by Chalfonte and Johnson (1996). Younger and older participants viewed line drawings of everyday objects presented in a grid and were asked to learn only the items, only the location of the items, only the colour of the items, or a combination of these features. In subsequent tests of recognition, older adults performed as well as younger adults when they had learned item-only and were tested on item-only, and when they had learned colour-only and were tested on colour-only (note that a deficiency in the location-only test was reported). Yet, when older adults had learned item *and* colour and they were subsequently tested on item *and* colour, their corrected recognition scores were ~30% worse than their younger counterparts. Therefore, older adults in this study demonstrated intact memory for item and study context (colour) independently, yet, they were impaired when the task required them to bind item and study context into a single complex unit.

Naveh-Benjamin (2000) extended the work of Chalfonte and Johnson (1996) with his associative deficit hypothesis. The associative deficit hypothesis states that older adults are less able to create and retrieve associative relationships between single units of information. The units might include an item and its study context, two items, or two study contexts. In a series of experiments, younger and older participants' learned a range of dual-unit stimuli including: word-nonword pairs (Exp 1; e.g., *lettuce-spink, castle-jown*), word-word pairs that were either semantically related (Exp 4) or semantically unrelated (Exps 2 and 4), and word-font combinations (Exp 3). The general procedure across experiments was for participants to study the stimuli and then complete three tests: two tests of item recognition (e.g., Exp 1 – selecting studied words from a group of distractors, i.e., *lettuce*; and selecting studied nonwords from a group of distractors, i.e., *spink*) and one test of associative recognition (e.g., Exp 1 – selecting studied word-nonword pairs from a group of recombined distractors, i.e., *lettuce-spink* but not *castle-spink*). Naveh-Benjamin supported his hypothesis by repeatedly showing that older adults' corrected recognition performance on the two item recognition tests was equivalent to younger adults; however, their corrected recognition performance on the associative recognition test was significantly worse than that of the younger adults.

The associative deficit hypothesis is also supported by neuroimaging evidence. Regions of hippocampus and prefrontal cortex have been shown to mediate successful associative binding and activation in these areas depends on age. In one study, younger and older participants were scanned while completing a series of working memory tasks (memory for objects, locations, and objects *and* locations). For the binding condition, older adults showed significantly less activity in left anterior hippocampus and medial frontal gyrus compared to younger adults (Mitchell, Johnson, Raye, & D’Esposito, 2000). Relatedly, the same regions associated with binding show preferential grey and white matter volume loss in aging (e.g., Raz et al., 2005). Thus, with increasing age, brain regions that mediate the associative binding process may become functionally and structurally weaker. In summary, behavioural and neural evidence conclude that older adults are less successful at associating to-be-learned items with contextual information during study compared to younger adults.

Goals and Hypotheses of the Current Study

The effect of study context on the distributed practice effect has been thoroughly investigated in young adult samples (Table 1). According to the hybrid account, a Lag × Study Context interaction occurs because massed, variably studied targets are assumed to benefit from complete reprocessing at S2 compared to massed, consistently studied targets and distributed, variably studied targets are assumed to benefit from additional encoding variability at S2, however, their study-phase retrieval is more likely to be impaired at S2 compared to distributed, consistently studied targets. Whether study context influences the distributed practice effect in aging is less clear. Based on prior evidence documenting older adults’ relative weakness in binding information about an item and its study context, it is possible that a change in study context may have less of an impact on the distributed practice effect in aging. In other words, varying study context between S1/S2 may not improve older adults’ memory for massed items and/or may not impair memory for distributed items. Consequently, older adults’ final memory performance, while influenced by repetition lag, should be less influenced by study context.

The current study sought to test the hypothesized Age × Lag × Study Context interaction. Younger and older participants completed a distributed practice paradigm where items were studied once or twice at lags of 0, 6, or 12 intervening items in either a consistent background scene or a variable background scene. On each trial, participants judged whether an item was new/old (item recognition) and, if it was old, whether the background scene remained the same or changed since the item's first presentation (study context recognition). After the study phase, participants completed a free recall test by writing down all of the items they remembered. Therefore, the experiment was a 2 (Age: younger or older; between-subjects) × 3 (Lag: 0, 6, or 12 intervening items; within-subjects) × 2 (Study Context: consistent or variable; within-subjects) mixed factorial design. The main dependent measure of interest was percentage of items correctly recalled on the final test. As a means of exploring hybrid account assumptions of the Lag × Study Context interaction, participants' responses at S1/S2 were also recorded for recognition accuracy.

The following a priori hypotheses were made:

Hypothesis 1: At S2, older adults should have comparable overall item recognition to younger adults; however, their overall study context recognition should be impaired. These findings would provide evidence in support of the associative deficit hypothesis in aging.

Hypothesis 2: An Age × Lag × Study Context interaction was expected in the study context recognition data at S2. Younger adults should accurately recognize the study context of massed targets—especially massed, variably studied targets—and they should be relatively impaired at recognizing the study context of distributed targets—

especially distributed, variably studied targets. These two findings would provide evidence in support of predictions made by the hybrid account[7]. Older adults should accurately recognize the study context of massed targets and they should be relatively impaired at recognizing the study context of distributed targets; however, manipulating study context should not influence older adults' study context recognition differently across lag conditions.

Hypothesis 3: An Age × Lag × Study Context interaction was expected at final test. Younger adults should recall a greater number of distributed targets compared to massed targets, but only if repeated in a consistent study context at S2. Older adults should recall a greater number of distributed targets, regardless of their study context at S2.

Experiment 2A

Method

Participants. Prior to recruiting participants, a power analysis was conducted to estimate the sample size required to detect an Age × Lag × Study Context interaction using a 95% power criterion. Assuming an estimated average *d* of 0.57 (based on various related effect sizes taken from Balota et al., 1989; Bercovitz et al. 2017; Chalfonte & Johnson, 1996; Naveh-Benjamin, 2000; Simone et al., 2012), a total sample size of at least 36 participants was recommended (18 participants per age group).

---

[7] A complementary hypothesis was also proposed. Younger adults' may benefit from variable study at a distributed lag so long as the lag is optimally timed. If this prediction were supported by the data, younger adults should accurately recognize the study context of Lag 0 and Lag 6 targets (but not Lag 12 targets).

Individuals were eligible to participate in the experiment if they were fluent in English, had normal or corrected-to-normal vision, and they passed a colourblindess test. Older adults were also screened for medical conditions that might impact cognitive ability. Specific ineligibility criteria included: hypertension, diabetes, anxiety and depression (unless any of these conditions was controlled by medication and/or behavioural therapy), as well as history of stroke or head injury.

The younger adult sample consisted of 36 York University students recruited from the Undergraduate Research Participant Pool. In exchange for participating in the experiment, they received course credit. Data from three younger adult participants were discarded. One participant was outside of the preferred age range (she was 39 years old) and two participants had abnormally low item recognition scores at S2 (e.g., 0% correct for massed items). Therefore, the final younger adult sample consisted of 33 participants (24 female; $M_{\text{age}} = 20.76$, $SD_{\text{age}} = 2.74$, range: 18-30 years).

The older adult sample consisted of 38 adults recruited from the Living and Learning in Retirement Group. The Living and Learning in Retirement Group is a collection of retired adults who attend weekly lectures on various topics hosted by Glendon College (York University). In exchange for participating in the experiment, they received monetary compensation. Data from four older adult participants were discarded. The sessions of two participants were interrupted by fire alarms. Two other participants completed the paradigm too slowly (i.e., >25% of their responses could not be recorded by the computer). Therefore, the final older adult sample consisted of 34 participants (23 female; $M_{\text{age}} = 72.24$, $SD_{\text{age}} = 7.36$, range: 60-87 years). A complete profile of the two samples is described in Table 3. The experiment was approved by York University's

Research Ethics Board and written consent was obtained from all participants prior to beginning the session.

Table 3

Experiment 2A: Sample Characteristics

Variable	Young Adults (*n* = 33, 24 female)		Older Adults (*n* = 34, 23 female)		
	M	*SD*	*M*	*SD*	*p*
Age	20.76	2.74	72.24	7.36	< .001
Years of Education	13.82	1.16	16.65	2.39	< .001
Shipley Vocabulary	95.91	11.78	107.59	8.53	< .001
Shipley Abstraction	104.76	12.47	109.76	9.12	.07
Backward Digit Span	6.85	2.46	7.38	2.16	.34
MoCA	--	--	26.97	2.11	--

Notes. Twelve years of education is equivalent to achieving a high school diploma. Shipley scores are standardized using lifespan norms. Max. scores for the backward digit span and the MoCA (Montreal Cognitive Assessment) are 14 and 30, respectively.

Materials (Experimental paradigm). The following materials were used to create the distributed practice paradigm.

Objects and background scenes. One hundred and five colour images of everyday objects were downloaded from the Aging Mind Laboratory's Object/Scene Database (http://agingmind.utdallas.edu/). The Object/Scene Database is an open-source tool for experimental research; it is used widely in studies of memory processing (e.g., Chee et al., 2006; Goh et al., 2004). The stimuli include both living and non-living objects from a variety of semantic categories (e.g., animals, locomotion, sporting goods, food, clothing, electronics, tools). Objects were chosen as the to-be-learned stimuli for two reasons. First, because free recall was the main dependent measure in the experiment, and free recall has been shown to be particularly affected by aging (compared to, for example, cued recall and/or recognition; e.g., Balota, Dolan, & Duchek, 2000; Grady & Craik, 2000), there was concern that older adults would perform at floor in a word learning paradigm. To avoid floor effects, a set of stimuli that could be satisfactorily learned by differently aged participants was needed. Objects can be considered a type of hybrid stimuli because they contain both verbal (label) and non-verbal (image) elements. Paivio and Csapo (1973) first demonstrated that objects are remembered better than words likely as a result of a dual-coding mechanism. Subsequent research has confirmed this finding, known as the pictorial superiority effect, in older adults (Park, Puglisi, & Sovacool, 1983). Therefore, by using object stimuli, floor effects in the older adult sample could be avoided. A second reason for using object stimuli was to improve the ecological validity of previous study context manipulations reported in the literature. Unlike artificial changes to the font or colour of words, study context manipulations applied to objects were more likely to reflect how these items are actually experienced in the real world. Therefore, by using object stimuli, the Lag × Study Context interaction finding could be tested under more authentic learning conditions. Although not replicated as many times as the verbal learning literature, a Lag × Study Context interaction has been reported for object stimuli learned by younger adults (Paivio, 1974; von Wright, 1976).

In order for participants to complete the experiment without becoming fatigued, and to avoid general floor effects associated with lengthy free recall tasks, the experimental paradigm was split into three blocks. Accordingly, the 105 objects were divided into three separate groupings, which were not biased toward any particular semantic category. Each grouping consisted of 35 objects. Twenty-four objects were repeated targets, with four objects assigned to each of the six experimental conditions (Lag 0, consistent study; Lag 0, variable study; Lag 6, consistent study; Lag 6, variable study; Lag 12, consistent study; Lag 12, variable study). The target "quadruples" were counterbalanced across conditions using a Latin Square design to ensure that each target object appeared in each of the six experimental conditions across participants. Five objects were once-presented non-targets and six objects were once-presented primacy buffers (*n* = 3) and recency buffers (*n* = 3). Thus, across the three blocks, there were 72 target objects (12 per condition), 15 once-presented non-targets, and 18 once-presented buffers.

Two background scenes were downloaded from the Object/Scene Database: a cityscape and a forest landscape. Using Adobe Photoshop, each target object was placed in each of the two background scenes and each non-target object (i.e., once-presented items and buffers) was randomly placed in either the cityscape or forest landscape. The position of an object within its scene at S1 was exactly the same position of the object within its scene at S2. However, across all trials, the positions of the objects varied. The purpose of varying objects' positions was to encourage more complete encoding of the scenes compared to if every object appeared consistently at a center fixation. In this way, the study context manipulation should be more successful. After image editing, there were a total of 177 object/scene image files.

Stimulus presentation. To ensure that targets across all Lag × Study Context conditions were evenly presented throughout a block, a custom programming script was written in MATLAB. This control measure, which serves to equate the retention interval between massed and distributed repetitions, is especially important in paradigms that end in free recall. As illustrated in Figure 1, the retention interval refers to the amount of time between an item's second presentation (S2) and the start of a free recall test. Given the nature of distributed repetition, it is more likely that S2 presentations of distributed items will appear later in a study list compared to S2 presentations of massed items. Consequently, the retention interval for distributed items may be shorter than the retention interval for massed items and any recall advantage associated with distributed practice may actually be an artifact of the recency effect, an argument first raised by Underwood (1969).

The MATLAB script began as an empty array consisting of 59 positions[8]. First, positions for the primacy and recency buffers were assigned. Second, the S2 presentation of each target in each of the six Lag × Study Context conditions was distributed evenly throughout the block such that the average S2 position for targets of each condition was statistically equivalent. Third, the corresponding S1 positions were filled. Finally, any positions that were still empty were filled with once-presented items. The number of once-presented items was constrained to be as few as possible so that the study list did not become too long. The MATLAB script successfully created three stimulus order outputs, one for each block of the experiment.

Using the three order outputs and the object/scene images, the experiment was programmed in E-Prime. Background order (between S1/S2) was counterbalanced across trials. Of the four target objects in the Lag 0, consistent study condition in a given block,

---

[8] Per experimental condition, there were four targets that each appeared two times. There were also five non-targets that each appeared once, three primacy buffers that each appeared once, and three recency buffers that each appeared once. Therefore, the total number of positions in a given block was 59. $[(6(4*2)) + 5 + 3 + 3 = 59]$.

two appeared in the cityscape at S1 and S2 and two appeared in the forest landscape at S1 and S2. Of the four target objects in the Lag 0, variable study condition in a given block, two appeared in the cityscape at S1 and the forest landscape at S2 and two appeared in the forest landscape in S1 and the cityscape at S2. Of the four target objects in the Lag 6, consistent study condition in a given block, two appeared in the cityscape at S1 and S2 and two appeared in the forest landscape at S1 and S2. Of the four target objects in the Lag 6, variable study condition in a given block, two appeared in the cityscape at S1 and the forest landscape at S2 and two appeared in the forest landscape at S1 and the cityscape at S2. Of the four target objects in the Lag 12, consistent study condition in a given block, two appeared in the cityscape at S1 and S2 and two appeared in the forest landscape at S1 and S2. Of the four target objects in the Lag 12, variable study condition in a given block, two appeared in the cityscape at S1 and the forest landscape at S2 and two appeared in the forest landscape at S1 and the cityscape at S2. As previously detailed, the five non-target objects and six buffers in a given block were randomly assigned to appear in the cityscape or forest landscape backgrounds.

Materials (Participant data). The following materials were used to gather information about the participant sample.

Colour vision test. The Pseudoisochromatic Plates Ishihara Compatible Color Vision Test was used to screen participants for red-green (RG) and yellow-blue (YB) colour perception deficiencies (Waggoner, 2005). The test has 17 plates (14 RG and 3 YB). Each plate illustrates a single- or double-digit number using dots of various sizes and colours. The participant is instructed to read aloud the number that he/she sees. The test is scored using a pass/fail method.

Shipley-2. The Shipley-2 provides an estimate of an individual's general cognitive functioning (Shipley, Gruber, Martin, & Klein, 2009). For the present experiment, the Vocabulary and Abstraction subscales were used, which measure crystalized intelligence and fluid intelligence, respectively. In the Vocabulary subscale, the participant chooses a word (from a series of four words) that is closest in meaning to a target word. There are 40 questions of increasing difficulty. In the Abstraction subscale, the participant fills in the missing item of a pattern (i.e., a word, letter, or number that completes a sequence). There are 25 questions of increasing difficulty. Raw scores for each of the subscales are tallied and converted to standardized scores using lifespan norms. The Shipley-2 has high reliability with an adult sample (alphas ranging from .88 to .97; Shipley et al., 2009).

Backward digit span test. The backward digit span test is a measure of working memory from the Wechsler Adult Intelligence Scale (Wechsler, 2008). Working memory refers to an individual's ability to retain and manipulate information, typically over the course of several seconds. On each trial of the test, the participant listens to a series of digits presented aloud, one-at-a-time, at a rate of 1 s each. The series may contain anywhere from two to eight digits. The participant immediately repeats the series back to the experimenter in reverse order. There are 14 trials of increasing difficulty (i.e., more numbers added to the series), organized into 7 sets of 2 trials each. The test is terminated when a participant fails to correctly answer both trials of a given set.

Montreal Cognitive Assessment. The Montreal Cognitive Assessment (MoCA) is a screening tool designed to assist physicians in the detection of early cognitive impairment in older adults (Nasreddine et al., 2005). The test is sensitive to frontally-mediated executive functioning and attention abilities, which makes it a strong tool in detecting mild cognitive impairment, stronger than the more traditional but less sensitive Mini Mental Status Exam (Smith, Gildeh, & Holmes, 2007). A broad range of cognitive domains are evaluated in the assessment, including: visuospatial ability (e.g., trail-making, copying a cube, drawing a clock), working memory, short-term memory, and long-term memory (e.g., digit span, word recall), attention (e.g., finger tapping), verbal fluency (e.g., sentence repetition, animal naming, words beginning with F), abstract thinking (e.g., identifying similar attributes between two items), and orientation (e.g., stating the current date and location). The test is administered by the experimenter and scored out of 30 points. According to creators of the test, a score of 26 is considered normal; however, a recent meta-analysis advises researchers to use a cut-off score of 23 to avoid false-positive diagnoses of cognitive impairment (Carson, Leach, & Murphy, 2018). All older adult participants in the experiment attained a score of 23 or higher.

Procedure. After providing consent to participate in the experiment, participants completed the colour vision test and the backward digit span test. Next, they completed a practice block of the experimental paradigm. Participants were seated approximately 50 cm from a 24” LCD computer monitor. The experimenter explained the nature of the object/scene images and encouraged participants to mentally link the objects and their background scenes in order to successfully complete the task. Participants were also aware that the experimenter was interested in memory for visual information and to expect a memory test later in the session.

In the experimental paradigm, the object/scene images appeared one-at-a-time in the centre of the computer screen at a rate of 4 s each. The question “OLD or NEW?” appeared above the stimulus image. For each image, participants were instructed to respond (by pressing one of two keys on the keyboard) whether they felt the *object* was new or old to them. If they decided the object was new, they were directed to a black screen for 4 s. If they decided the object was old, they were directed to a black screen with the following statement for 4 s: “You said old. SAME or DIFFERENT background?” They were instructed to respond (by pressing one of two other keys on the keyboard) whether the object’s *background scene* was the same or different than its first presentation. Therefore, each trial was 8 s in duration (4 s for an old/new response and 4 s for either a black screen or a same/different response), followed by a 500 ms fixation cross between trials. Responses were recorded for accuracy. Handedness of response keys was counterbalanced across participants. At the end of the study phase, the experimenter talked with the participant about his/her performance on the task for approximately one minute. Finally, participants were given three minutes to write down all of the objects they could remember from the study phase. When the time limit expired, the experimenter reviewed the list of recalled objects and clarified any object that was obscurely labelled. Figure 9 provides an illustration of the paradigm.

After completing the first block of the experimental paradigm, participants completed the Shipley Vocabulary subscale. Next, they completed the second block of the paradigm and the Shipley Abstraction subscale. After a break, they completed the third block of the paradigm and older adults completed the Montreal Cognitive Assessment.

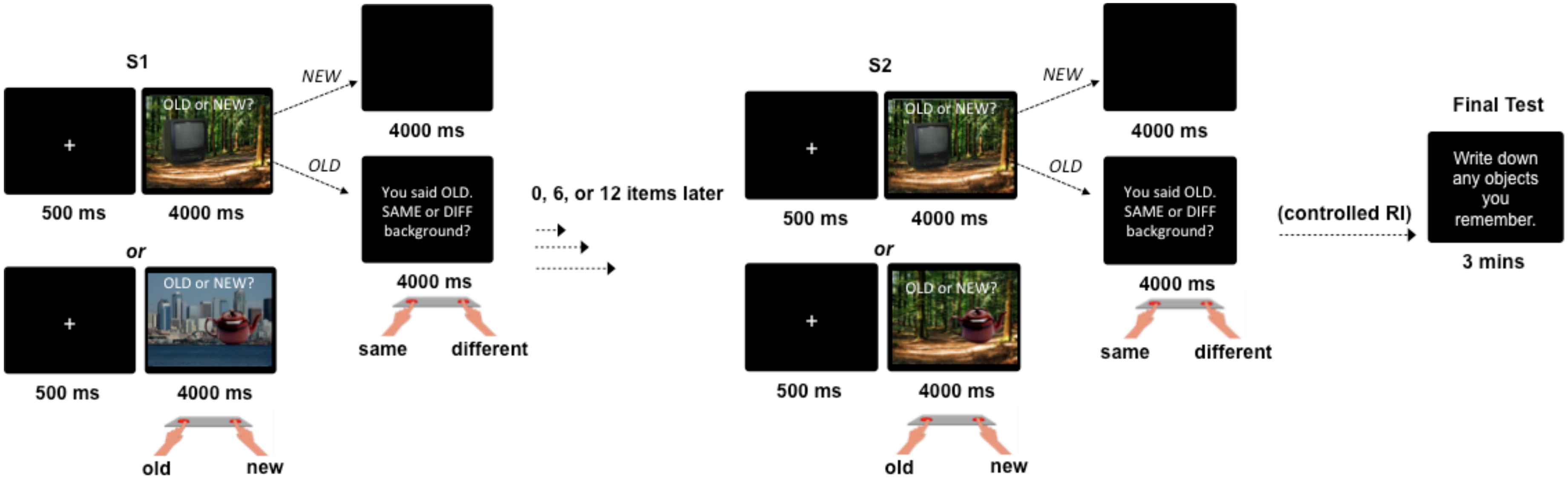

Figure 9. The distributed practice paradigm used in Experiment 2A. Objects (e.g., television, teapot) were presented for study (S1) and repeated for study (S2) after a lag of 0, 6, or 12 intervening items in either the same background scene (consistent study context) or a different background scene (variable study context) as their first presentations. The background scenes were a cityscape or a forest landscape. On each trial, participants judged whether the object was new/old (item recognition) and, if it was old, whether the background scene was the same or different from its initial presentation (study context recognition). After the study phase, participants were given three minutes to recall all of the objects.

Results

Data from the study phase includes participants' item recognition accuracy at S1, item recognition accuracy at S2, and study context recognition accuracy at S2 (for correctly recognized items)[9]. Recognition accuracy is reported as percentage correct. Correct recognition implies that participants made a correct response in the allotted amount of time (4 s). Incorrect recognition implies that participants made an erroneous response or did not reply in the allotted amount of time (miss)[10]. Data from the final test includes percentage of items correctly recalled. Unless otherwise stated, the specific analysis conducted was a 2 × 3 × 2 mixed factorial ANOVA. Bayes factors were also calculated to confirm main effects and interactions and clarify any marginally significant results.

Study Phase Data

Data for item recognition accuracy at S2 and study context recognition accuracy at S2 are reported in Table 5 and Table 6, respectively. The data were left-skewed according to visual inspection of histograms and a series of significant Shapiro-Wilk tests. Despite non-normality, analyses were conducted using parametric statistics, which are known to be robust to violations of normality given an adequate sample size (according to central limit theorem, $n \geq 30$ per group; Field, 2009). They are the preferred analysis technique over non-parametric tests due to greater statistical power.

---

[9] For item recognition at S1, performance was at ceiling (Table 4). Since these data were not the foci of specific hypotheses, they were not formally analyzed.

[10] Misses were relatively rare. Item recognition misses, collapsed across condition, were very low (for younger adults: $M = 1.08$, $SD = 1.34$; for older adults: $M = 2.86$, $SD = 3.16$). Study context recognition misses, collapsed across condition, were also very low (for younger adults $M = 0.69$, $SD = 0.96$; for older adults: $M = 3.40$, $SD = 4.42$).

Item recognition at S2. The only notable finding from this analysis was a significant Lag × Study Context interaction, $F(2, 130) = 7.18$, $p = .001$, $\eta_p^2 = .10$, $BF_{Inclusion} = 0.93$ (anecdotal evidence for H_0), suggesting that variable study particularly impaired recognition of Lag 12 items but not Lag 0 or Lag 6 items. Since the Bayes factor opposed the effect, it was not considered further. Note that the main effect of Age was not significant; item recognition accuracy at S2 was similar for younger adults ($M = 95.75$, $SD = 6.13$) and older adults ($M = 93.59$, $SD = 5.20$).

Study context recognition at S2. This analysis revealed three findings. First, there was a significant main effect of Age, $F(1, 65) = 7.89$, $p < .01$, $d = 0.70$, $BF_{Inclusion} = 2.12$ (anecdotal evidence for H_1). Younger adults had superior study context recognition accuracy ($M = 83.89$, $SD = 8.32$) compared to older adults ($M = 77.04$, $SD = 11.35$). Second, there was a significant main effect of Lag, $F(1.60, 104.15) = 14.87$, $p < .001$, $\eta_p^2 = .19$, $BF_{Inclusion} = 31.84$ (strong evidence for H_1). Post-hoc, Bonferroni-corrected pairwise comparisons showed that study context recognition of Lag 0 items ($M = 84.78$, $SD = 11.71$) was superior to study context recognition of both Lag 6 items ($M = 79.12$, $SD = 11.84$) and Lag 12 items ($M = 77.34$, $SD = 13.41$; $ps < .001$); however, Lag 6 and Lag 12 items did not significantly differ from each other. Third, there was a significant main effect of Study Context, $F(1, 65) = 85.86$, $p < .001$, $d = 1.24$, $BF_{Inclusion} = \infty$ (extreme evidence for H_1). Study context recognition of consistently studied items ($M = 91.64$, $SD = 8.31$) was superior to study context recognition of variably studied items ($M = 69.19$, $SD = 18.39$). Finally, there was a marginally significant Lag × Study Context interaction, $F(1.64, 106.31) = 3.02$, $p = .06$, $\eta_p^2 = .04$, $BF_{Inclusion} = 0.85$ (anecdotal evidence for H_0), a trend which suggested that variable study impaired study context

Table 4

Experiment 2A: Item Recognition at S1 (% Correct)

Age Group	Lag 0 Consistent		Lag 0 Variable		Lag 6 Consistent		Lag 6 Variable		Lag 12 Consistent		Lag 12 Variable		Non-Targets	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Young Adults	98.48	4.39	98.74	3.68	97.73	5.62	99.24	2.43	97.48	4.88	97.22	5.38	97.37	4.06
Older Adults	95.59	7.18	97.55	4.82	97.55	7.26	98.78	3.00	98.78	3.00	97.55	4.37	92.55	8.00

Table 5

Experiment 2A: Item Recognition at S2 (% Correct)

Age Group	Lag 0 Consistent		Lag 0 Variable		Lag 6 Consistent		Lag 6 Variable		Lag 12 Consistent		Lag 12 Variable	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Young Adults	96.47	5.52	96.47	5.52	95.96	7.25	96.47	6.60	96.21	10.64	92.93	13.03
Older Adults	92.16	9.39	93.63	8.22	95.10	6.84	94.85	6.80	95.59	7.46	90.20	10.95

Table 6

Experiment 2A: Study Context Recognition at S2 (% Correct)

Age Group	Lag 0 Consistent		Lag 0 Variable		Lag 6 Consistent		Lag 6 Variable		Lag 12 Consistent		Lag 12 Variable	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Young Adults	98.49	4.87	78.61	18.69	94.69	7.99	70.76	20.01	90.98	7.88	69.81	24.30
Older Adults	89.22	15.21	73.04	21.01	89.69	10.85	61.57	23.27	87.05	13.14	61.71	25.37

recognition at each lag but especially at the two distributed lags. Since the Bayes factor opposed the interaction trend, it was not considered further.

Final Test Data

The final test data are illustrated in Figures 10a (younger adults) and 10b (older adults). The only notable finding from this analysis was a significant main effect of Lag, $F(2, 130) = 134.48$, $p < .001$, $\eta_p^2 = .67$, $\text{BF}_{\text{Inclusion}} = 3.21 \times 10^{15}$ (i.e., approaching ∞; extreme evidence for H_1). Post-hoc, Bonferroni-corrected, pairwise comparisons showed that recall of Lag 6 items ($M = 46.14$, $SD = 13.69$) and Lag 12 items ($M = 48.69$, $SD = 13.71$) was superior to recall of Lag 0 items ($M = 25.44$, $SD = 12.57$; ps $< .001$); however, recall of Lag 6 and Lag 12 targets did not significantly differ from each other. Note that the Age × Lag × Study Context interaction was not significant.

Discussion

The main goal of Experiment 2A was to test the hypothesis that study context influences the distributed practice effect for younger adults but not older adults. This hypothesis was derived from research documenting older adults' relative impairment in remembering contextual features associated with an item versus the item itself (Spencer & Raz, 1995) as well as difficulties binding item and study context into a single memory trace (Naveh-Benjamin, 2000). Older adults were predicted to behave differently from younger adults during the study phase, which would, in turn, drive differences in final recall performance between the two age groups. More specifically, the factor of lag was expected to be the primary influence on older adults' S2 recognition responses whereas the interaction of lag and study context was expected to influence younger adults' S2

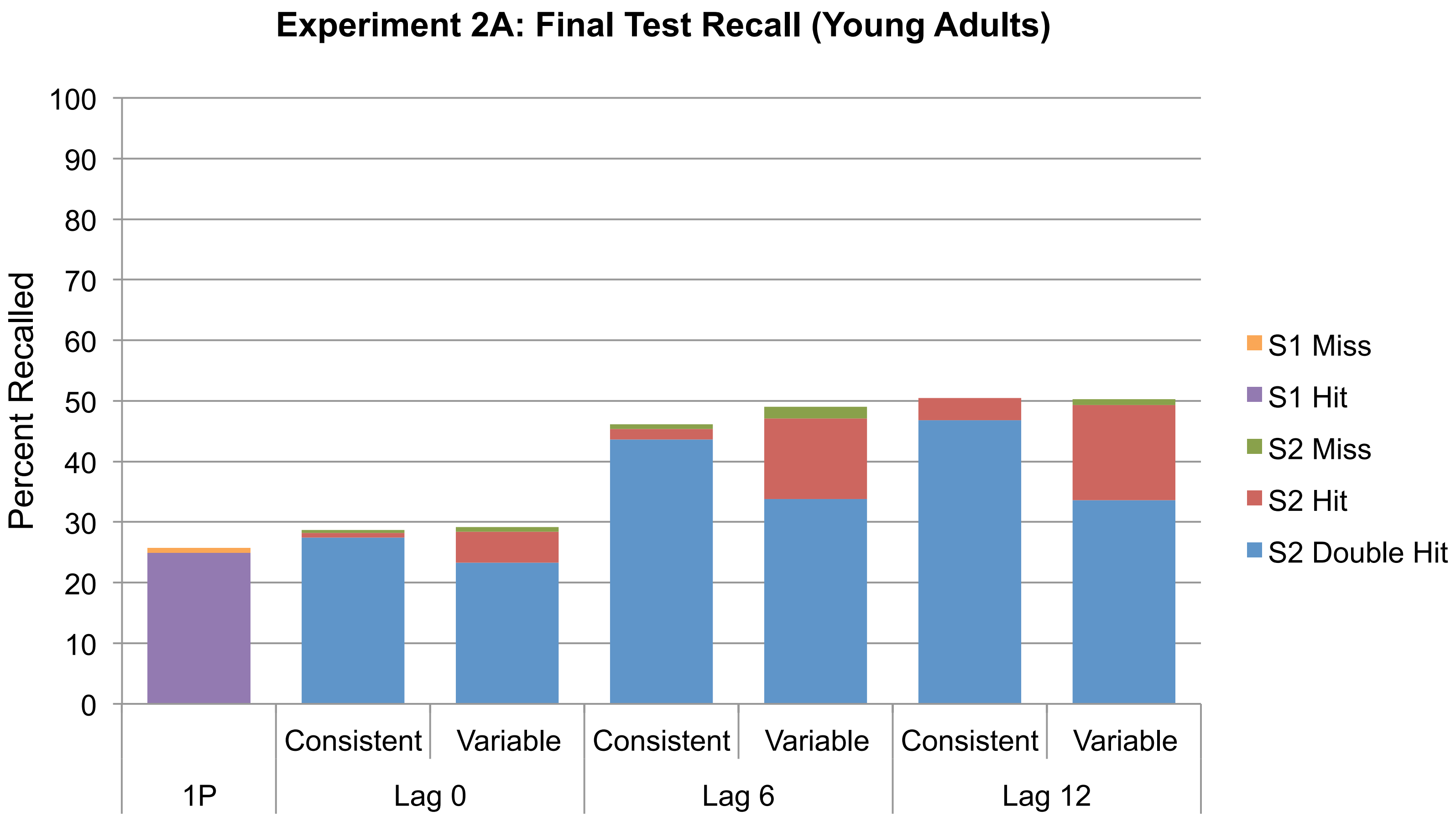

Figure 10a. Younger adults' percentage of correctly recalled objects on the final test. "Hit" represents items that were correctly recognized as repeats at S2. "Miss" represents items that were not recognized as repeats at S2. "Double Hit" represents items that were correctly recognized as repeats *and* correctly recognized for their study context at S2.

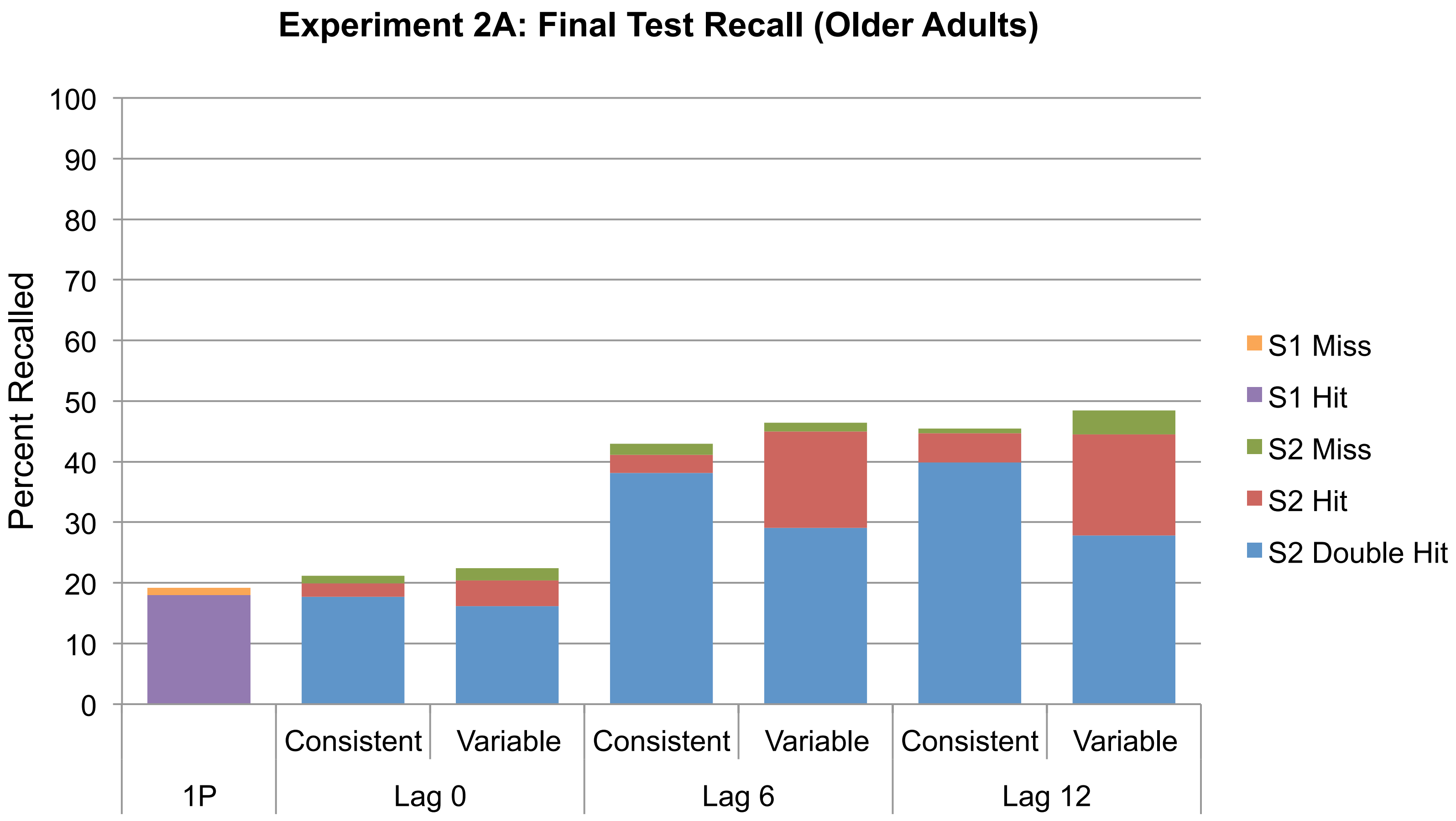

Figure 10b. Older adults' percentage of correctly recalled objects on the final test. "Hit" represents items that were correctly recognized as repeats at S2. "Miss" represents items that were not recognized as repeats at S2. "Double Hit" represents items that were correctly recognized as repeats *and* correctly recognized for their study context at S2.

recognition responses. Contrary to predictions, younger and older participants exhibited similar patterns of performance at S2 and the Age × Lag × Study Context interaction in object free recall was not significant. The Lag × Study Context interaction with younger participants was not also replicated. Failure to replicate this finding is discussed below.

Age Differences in Item Memory vs. Study Context Memory

Hypothesis 1 predicted different effects of age on item memory versus study context memory at S2. Supporting this prediction, older adults' overall item recognition at S2 was similar to that of younger adults (94% vs. 96%, respectively); however, their overall study context recognition was impaired compared to younger adults (77% vs. 84%, respectively). These results converge with previous research documenting older adults' difficulty retrieving information about an item's contextual features during study (Spencer & Raz, 1995). That being said, older adults exhibited the same pattern of behaviour as younger adults in terms of relative impairments in item recognition and study context recognition of distributed, variably studied items compared to distributed, consistently studied items (Tables 5 and 6). This finding is somewhat challenging for the associative deficit hypothesis, which states that older adults are less likely to bind an item and its contextual features during study (Naveh-Benjamin, 2000). With respect to item recognition, if older adults were less likely to bind item and study context at S1, it should follow that their item recognition scores at S2 for consistently studied targets versus variably studied targets should be equivalent. In other words, if an item was never bound to a study context at S1, the item itself should be recognizable in the (near) future regardless of its future study context. Since older adults' item recognition of Lag 12, variably studied items was worse than their item recognition of Lag 12, consistently studied items, older adults must have bound item and study context at S1, which caused interference in item recognition when the item reappeared at S2 in a different background scene, just as it did for younger adults. With respect to study context recognition, if older adults were less likely to bind item and study context at S1, it should follow that their study context recognition scores at S2 for consistently studied items versus variably studied items should also be equivalent. In other words, if an item was never bound to a study context at S1, and an evaluation about the item's study context is made at S2, the chances of a correct study context recognition response for consistently studied items versus variably studied items should be about the same (theoretically, a 50/50 chance of making a correct response). Since older adults' study context recognition of distributed, variably studied items was worse than their study context recognition of distributed, consistently studied items, older adults must have bound item and study context at S1, making it more difficult to evaluate the item's study context when it reappeared at S2 in a different form, just like the difficulty experienced by younger adults. Thus, although older adults generally remembered less about an item's prior study context compared to younger adults, there is evidence that older adults engaged in some degree of associative binding at S1 that subsequently influenced their S2 performance in a pattern very similar to that of younger adults.

The Effects of Lag and Study Context During the Study Phase

Hypothesis 2 predicted independent and interactive effects of lag and study context on S2 performance and further predicted that these effects might be different depending on age. In general, manipulations of lag and study context had minimal effects on participants' item recognition at S2. Item recognition was very high for all participants across all conditions (on average, >90%; Table 5). Correct item recognition implies that participants engaged in at least a rudimentary form of study-phase retrieval, a process analogous to successful familiarity (Yonelinas, 2002). Only when an item reappeared in a different background scene at the most distributed lag did performance begin to suffer, for both younger and older adults. Although this finding was not statistically robust according to Bayesian statistics, and performance was still very good in this condition, it does suggest that study-phase retrieval of Lag 12, variably studied items was most likely to fail relative to all other repeated items.

Lag and study context had more noticeable effects on participants' study context recognition at S2, which was the focus of Hypothesis 2. Correct study context recognition implies that participants engaged in a deeper form of study-phase retrieval, a process analogous to successful recollection. Only by recollection can a memory trace reap the benefits of S1 retrieval and trace updating (Yonelinas, 2002).

Younger adults' study context recognition at S2 provided mixed support for Hypothesis 2. On the one hand, younger adults' study context recognition of Lag 0, variably studied items was ~20% worse than Lag 0, consistently studied items, which was a pattern of results opposite to those predicted. If variation in study context counteracts typical deficient processing associated with massed repetition, it was expected that study context recognition of Lag 0, variably studied items should be nearly perfect (and/or perhaps surpass recognition of Lag 0, consistently studied items, which were expected to elicit a greater number of errors due to inattention). It is possible, however, that the opposite logic was true. Participants may have been less certain about their study context recognition responses to Lag 0, variably studied items compared to

Lag 0, consistently studied items and these errors may have caused greater reflective processing of Lag 0, variably studied items. Unfortunately, the S2 data cannot clearly speak to either interpretation. On the other hand, younger adults exhibited larger study context recognition impairments for variably studied targets at the two distributed lags (Lag 6 and Lag 12) compared to consistently studied targets, which was a predicted pattern of results. Although this finding was not statistically robust according to Bayesian statistics, it does suggest that study-phase retrieval of distributed, variably studied items was least likely to be successful. Taken together, the results demonstrate that younger adults' study context recognition significantly deteriorated with increasing lag; however, the influence of study context on lag was not exactly the crossover interaction that was hypothesized[11].

Older adults' study context recognition at S2 also provided mixed support for Hypothesis 2. On the one hand, older adults' study context recognition deteriorated with increasing lag, which was a predicted pattern of results. On the other hand, older adults exhibited the same Lag × Study Context interaction trend as younger adults, which was a pattern of results that was not expected. As discussed above, older adults did seem to engage in associative binding processes that were similar to younger adults (albeit weaker). Consequently, study context had a greater influence on S2 recognition responses for older adults than expected.

The Effects of Lag and Study Context at Final Test

---

[11] The additional prediction that younger adults may benefit from variable study at a distributed lag so long as the lag is optimally timed (in this case, Lag 6 but not Lag 12) was not supported by the S2 data.

Finally, Hypothesis 3 predicted that the Lag × Study Context interaction typically observed at final test would also depend on age. Contrary to expectations, the final test recall data showed that, on average, all participants exhibited a distributed practice effect. Items repeated at either of the two distributed lags (Lag 6 and Lag 12) were better recalled than items repeated at the massed lag (Lag 0). The effect was clear regardless of study context or age. Thus, although participants experienced greater study-phase retrieval difficulty at S2 for items repeated in variable study conditions versus consistent study conditions, final recall of items studied in the same background scene versus two different backgrounds scenes was equivalent. Furthermore, younger and older adults performed remarkably similar on the final test (see comparison of Figure 10a and 10b), which contradicts other research showing less of a distributed practice advantage for older adults compared to younger adults (e.g., Balota et al., 1989; Simone et al., 2012). Considering all results from the final test data, Hypothesis 3 was not supported.

Figures 10a and 10b illustrate participants' final test performance as a function of whether the items they recalled had been recognized as repetitions at S2 with or without subsequent study context recognition ("S2 Double Hit" and "S2 Hit", respectively). Most of the items that participants recalled on the final test were S2 Double Hit items. Interestingly, relative to recall of Lag 0, variably studied items, recall of Lag 6/12, variably studied items was more likely to include items that had been correctly recognized as repetitions at S2 but not correctly recognized for their study context. This finding reflects the fact that study-phase retrieval of distributed, variably studied items was most difficult for participants during the study phase. It may also suggest that distributed, variably studied items still underwent memory trace updating at S2 despite explicit study context recognition errors, which, in turn, increased probability of recalling these items on the final test. Conversely, massed, variably studied items that were correctly recognized as repetitions at S2 but not correctly recognized for their study context may have been less likely to engage in successful trace updating which, in turn, decreased probability of recalling these items on the final test.

The most unexpected result from this experiment was a failure to replicate the Lag × Study Context interaction for younger adults. Contrary to many previous studies, varying study context did not cause an increase in final recall of massed items nor did it cause a decrease in final recall of distributed items. Instead, younger adults exhibited a classic distributed practice effect regardless of whether the items were studied twice in the same background scene or twice in two different background scenes. One possible explanation for this finding relates to how variable study was operationally defined in the experiment. Specifically, variably studied items appeared in a total of only two background scenes during the study phase: a cityscape or a forest landscape. The background scenes were not uniquely associated with a given item; rather, they were shared among all items. A given item may have appeared in a different background scene at S2 compared to S1 (e.g., CITY-FOREST); however, this "different" background scene was not different from other repeated target objects. Relatedly, the background scenes were not semantically congruent with all of the objects. A considerable number of Lag × Study Context studies in the verbal learning domain have used study context manipulations that are unique to a given item. For example, participants might learn stimulus-response word pairs where the stimulus term is either the same or different at S1 and S2, followed by a final recall test of the response terms. In these studies, the stimulus terms are unique to a given response term and the terms are semantically associated (e.g., *speed-engine*; *valve-engine*; Madigan, 1969). It is possible that failure to replicate the Lag × Study Context interaction with younger adults in the present experiment was due to the limited number of background scenes, which may have inadvertently deflated the study context manipulation. To address this limitation, a second experiment was conducted. Experiment 2B is a replication of Experiment 2A using a new series of background scenes that were unique to each to-be-learned item. In every other respect, the methodology of Experiment 2B was the same as Experiment 2A.

Experiment 2B

Method

Participants. Participant eligibility criteria was the same as Experiment 2A. The younger adult sample consisted of 36 York University students recruited from the Undergraduate Research Participant Pool (29 female; $M_{\text{age}} = 19.75$, $SD_{\text{age}} = 2.64$, range: 17-30 years). In exchange for participating in the experiment, they received course credit.

The older adult sample consisted of 36 adults recruited from the York University Research Participant Pool. The York University Research Participant Pool is a collection of community members who are interested in participating in health research studies at the university and its affiliated institutions. In exchange for participating in the experiment, they received monetary compensation. Data from three participants were discarded. One participant was colourblind and two completed the experimental paradigm too slowly (i.e., >25% of their responses could not be recorded by the computer). Therefore, the final older adult sample consisted of 33 participants (18 female; $M_{\text{age}} = 71.45$, $SD_{\text{age}} = 5.12$; range: 61-81 years). A complete profile of the two samples is described in Table 7.

Table 7

Experiment 2B: Sample Characteristics

Variable	Young Adults (*n* = 36, 29 female)		Older Adults (*n* = 33, 18 female)		
	M	*SD*	*M*	*SD*	*p*
Age	19.75	2.64	71.45	5.12	< .001
Years of Education	12.97	1.07	17.34	3.21	< .001
Shipley Vocabulary	99.58	12.05	112.15	6.68	.05
Shipley Abstraction	100.92	10.89	108.88	12.01	< .01
Backward Digit Span	6.72	2.54	8.61	1.75	.42
MoCA	--	--	27.52	2.14	--

Notes. Twelve years of education is equivalent to achieving a high school diploma. Shipley scores are standardized using lifespan norms. Max. scores for the backward digit span and the MoCA (Montreal Cognitive Assessment) are 14 and 30, respectively.

Materials. The materials were the same as those used in Experiment 2A with the exception of the object/scene images. The objects were the same as Experiment 2A but, instead of a cityscape and a forest landscape, a new collection of background scenes was sourced. One hundred and seventy-seven different scenes were downloaded from the Aging Mind Laboratory's Object/Scene Database (http://agingmind.utdallas.edu/). The scenes were chosen based on their semantic congruency with each object stimulus. They were either interior scenes (e.g., a cubicle workspace, a hotel lobby) or exterior scenes (e.g., a country road, a building façade) and they did not contain any other prominent objects in the foreground. Using Adobe Photoshop, each target object was placed in each of two different and appropriate background scenes. Each non-target object (i.e., once-presented items and buffers) was placed in one appropriate background scene. Each object was placed in exactly the same position in the new scenes as it had appeared in the city/forest background scenes from Experiment 2A. For example, the object *piano* may have appeared in front of a city skyscraper in Experiment 2A and—in exactly the same position—in a living room in Experiment 2B. In the absence of eye tracking data, placing the objects in the same position between the two experiments provided some control over participants' encoding processes such that data from the two experiments could be compared. As in Experiment 2A, although the position of all objects varied across trials, a given repeated object appeared in the same position in its S1 scene and its S2 scene. After image editing, there were a total of 177 new object/scene files.

Procedure. The procedure was the same as Experiment 2A. Participants completed three blocks of the experimental paradigm as well as the colour vision test, backward digit span, Shipley-2, and the MoCA (older adults). Figure 11 provides an illustration of the paradigm.

Results

Study Phase Data

Data for item recognition accuracy at S1, item recognition accuracy at S2, and study context recognition accuracy at S2 (for correctly recognized items) are reported in Table 8, Table 9, and Table 10, respectively[12]/[13]. Despite non-normality in the study

---

[12] Once again, performance for item recognition at S1 was at ceiling.

[13] As in the previous experiment, correct recognition implies that participants made a correct response in the allotted amount of time (4 s). Incorrect recognition implies that participants made an erroneous response or did not reply in the allotted amount of time (miss). Once again, misses were relatively rare. Item recognition misses, collapsed across condition, were very low (for younger adults: *M* = 0.70, *SD* = 0.67; for older adults: *M* = 2.98, *SD* = 2.96). Context recognition misses, collapsed across condition, were also very low (for younger adults *M* = 0.61, *SD* = 1.04; for older adults: *M* = 2.39, *SD* = 2.96).

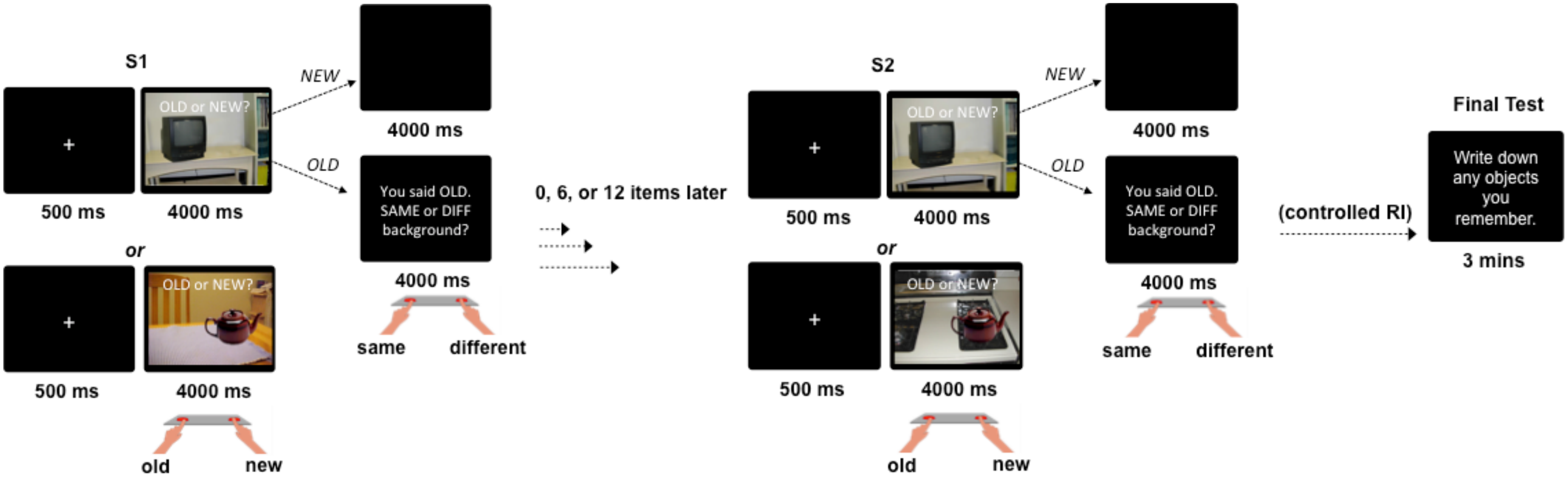

Figure 11. The distributed practice paradigm used in Experiment 2B. Objects (e.g., television, teapot) were presented for study (S1) and repeated for study (S2) after a lag of 0, 6, or 12 intervening items in either the same background scene (consistent study context) or a different background scene (variable study context) as their first presentations. The background scenes were unique to each to-be-learned object. On each trial, participants judged whether the object was new/old (item recognition) and, if it was old, whether the background scene was the same or different from its initial presentation (study context recognition). After the study phase, participants were given three minutes to recall all of the objects.

phase data, analyses were conducted using parametric statistics with the same justification as stated in Experiment 2A.

Item recognition at S2. This analysis revealed four findings. First, there was a significant main effect of Age, $F(1, 67) = 10.92$, $p < .01$, $d = 0.45$, $\text{BF}_{\text{Inclusion}} = 1959.25$ (extreme evidence for H_1). Younger adults recognized more repeated items ($M = 96.64$, $SD = 5.74$) than older adults ($M = 93.14$, $SD = 9.81$). Second, there was a significant main effect of Study Context, $F(1, 67) = 34.82$, $p < .001$, $d = 0.74$, $\text{BF}_{\text{Inclusion}} = 3.27 \times 10^9$ (extreme evidence for H_1). Recognition of consistently studied items ($M = 97.14$, $SD = 3.40$) was superior to recognition of variably studied items ($M = 92.79$, $SD = 7.48$). Third, there was a significant Age × Study Context interaction, $F(1, 67) = 12.64$, $p = .001$, $\eta_p^2 = .16$, $\text{BF}_{\text{Inclusion}} = 687.18$ (extreme evidence for H_1). Two separate post-hoc paired samples *t*-tests comparing item recognition accuracy between consistent and variable study conditions separately for each of the two age groups showed that, although variable study impaired item recognition for both age groups compared to consistent study, the effect size was larger for older adults [$t(32) = 5.36$, $p < .001$, $d = 1.12$, $\text{BF}_{10} = 2882.14$ (extreme evidence for H_1)] than it was for younger adults [$t(35) = 2.29$, $p < .05$, $d = 0.40$, $\text{BF}_{10} = 1.77$ (anecdotal evidence for H_1)]. Finally, there was a significant Lag × Study Context interaction, $F(2, 134) = 5.07$, $p < .01$, $\eta_p^2 = .07$, $\text{BF}_{\text{Inclusion}} = 0.44$ (anecdotal evidence for H_0), suggesting that suggesting that variable study particularly impaired recognition of Lag 6 and Lag 12 items but not Lag 0 items. Since the Bayes factor opposed the interaction effect, it was not considered further.

Study context recognition at S2. This analysis revealed six findings. First, there was a significant main effect of Age, $F(1, 67) = 16.56$, $p < .001$, $d = 0.99$, $\text{BF}_{\text{Inclusion}} =$

1.47×10^9 (extreme evidence for H_1). Younger adults had superior study context recognition accuracy ($M = 89.89$, $SD = 6.74$) than older adults ($M = 82.06$, $SD = 9.16$). Second, there was a significant main effect of Lag, $F(2, 134) = 6.27$, $p < .01$, $\eta_p^2 = .09$, $BF_{Inclusion} = 30.27$ (strong evidence for H_1). Post-hoc pairwise comparisons showed that study context recognition of Lag 0 items ($M = 88.47$, $SD = 10.50$) was superior to study context recognition of Lag 12 items ($M = 83.94$, $SD = 10.50$; $p < .01$); however, there were no differences between Lag 0 and Lag 6 items ($M = 86.01$, $SD = 11.23$) or Lag 6 and Lag 12 items. Third, there was a significant main effect of Study Context, $F(1, 67) = 77.60$, $p < .001$, $d = 0.96$, $BF_{Inclusion} = \infty$ (extreme evidence for H_1). Study context recognition of consistently studied items ($M = 92.91$, $SD = 7.37$) was superior to study context recognition of variably studied items ($M = 79.37$, $SD = 14.76$). Fourth, there was a significant Age × Study Context interaction, $F(1, 67) = 24.71$, $p < .001$, $\eta_p^2 = .27$, $BF_{Inclusion} = 6.33 \times 10^7$ (extreme evidence for H_1). Two post-hoc paired samples t-tests comparing study context recognition accuracy between consistent and variable study conditions separately for each of the two age groups showed that, although variable study impaired study context recognition for both age groups compared to consistent study, the effect size was larger for older adults [$t(32) = 8.68$, $p < .001$, $d = 1.69$, $BF_{10} = 1.73 \times 10^7$ (extreme evidence for H_1)] than it was for younger adults [$t(35) = 3.08$, $p < .01$, $d = 0.52$, $BF_{10} = 9.36$ (strong evidence for H_1)]. Fifth, there was a significant Lag × Study Context interaction, $F(2, 134) = 8.75$, $p < .001$, $\eta_p^2 = .12$, $BF_{Inclusion} = 59.73$ (very strong evidence for H_1). Three post-hoc paired samples t-tests comparing context recognition accuracy between consistent and variable study conditions separately for each of the three lags revealed that, although variable study impaired context recognition at all lags compared

Table 8

Experiment 2B: Item Recognition at S1 (% Correct)

Age Group	Lag 0 Consistent		Lag 0 Variable		Lag 6 Consistent		Lag 6 Variable		Lag 12 Consistent		Lag 12 Variable		Non-Targets	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Young Adults	97.46	3.89	97.22	4.45	99.07	2.66	97.69	5.50	97.92	4.62	99.31	2.34	98.21	3.96
Older Adults	95.20	7.23	97.48	4.41	97.22	5.38	97.98	4.18	96.72	5.08	96.97	5.83	95.24	7.07

Table 9

Experiment 2B: Item Recognition at S2 (% Correct)

Age Group	Lag 0 Consistent		Lag 0 Variable		Lag 6 Consistent		Lag 6 Variable		Lag 12 Consistent		Lag 12 Variable	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Young Adults	97.45	4.80	97.22	4.45	98.15	3.51	95.14	8.54	97.00	4.94	94.91	6.39
Older Adults	94.70	7.16	91.67	9.99	98.23	4.54	87.88	12.52	97.22	6.48	89.14	11.31

Table 10

Experiment 2B: Study Context Recognition at S2 (% Correct)

Age Group	Lag 0 Consistent		Lag 0 Variable		Lag 6 Consistent		Lag 6 Variable		Lag 12 Consistent		Lag 12 Variable	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Young Adults	93.70	8.02	90.53	13.00	92.95	9.51	88.13	11.79	92.09	10.88	81.93	17.53
Older Adults	90.29	10.70	78.72	18.69	94.12	10.74	68.02	20.37	94.33	7.85	66.86	18.72

to consistent study, the effect size was larger for Lag 6 items [$t(68) = 5.97, p < .001, d = 0.76$, $BF_{10} = 137415.81$ (extreme evidence for H_1)] and Lag 12 items [$t(68) = 6.83, p < .001, d = 0.86$, $BF_{10} = 3.98 \times 10^6$ (extreme evidence for H_1)] compared to Lag 0 items [$t(68) = 3.38, p = .001, d = 0.43$, $BF_{10} = 21.60$ (very strong evidence for H_1)]. Finally, there was a marginally significant Age × Lag × Study Context interaction, $F(2, 134) = 2.75, p = .07, \eta_p^2 = .04$, $BF_{Inclusion} = 0.87$ (anecdotal evidence for H_0). Since the Bayes factor opposed the interaction trend, it was not considered further.

Final Test Data

The final test data are illustrated in Figures 12a and 12b. In this analysis, there was a significant main effect of Lag, $F(2, 134) = 126.12, p < .001, \eta_p^2 = .65$, $BF_{Inclusion} = \infty$ (extreme evidence for H_1). Post-hoc, Bonferroni-corrected, pairwise comparisons showed that recall of Lag 6 items ($M = 54.77$, $SD = 13.75$) and Lag 12 targets ($M = 52.84$, $SD = 14.67$) was superior to recall of Lag 0 items ($M = 28.74$, $SD = 12.34$; ps < .001); however, recall of Lag 6 and Lag 12 items did not significantly differ from each other. There was also a significant main effect of Study Context, $F(1, 67) = 13.53, p < .001, \eta_p^2 = 0.47$, $BF_{Inclusion} = 67903.24$ (extreme evidence for H_1), indicating that recall of variably studied items ($M = 48.27$, $SD = 12.68$) was superior to recall of consistently studied items ($M = 42.63$, $SD = 11.61$). Finally, there was a significant Lag × Study Context interaction, $F(2, 134) = 14.11, p < .001, \eta_p^2 = .17$, $BF_{Inclusion} = 2985.86$ (extreme evidence for H_1). Three separate post-hoc paired samples *t*-tests comparing recall of consistently studied targets and variably studied targets separately for each of the three lags revealed that study context had a significant effect on final recall of Lag 6 items but not Lag 0 or Lag 12 items. Specifically, recall of Lag 6, variably studied items ($M =$

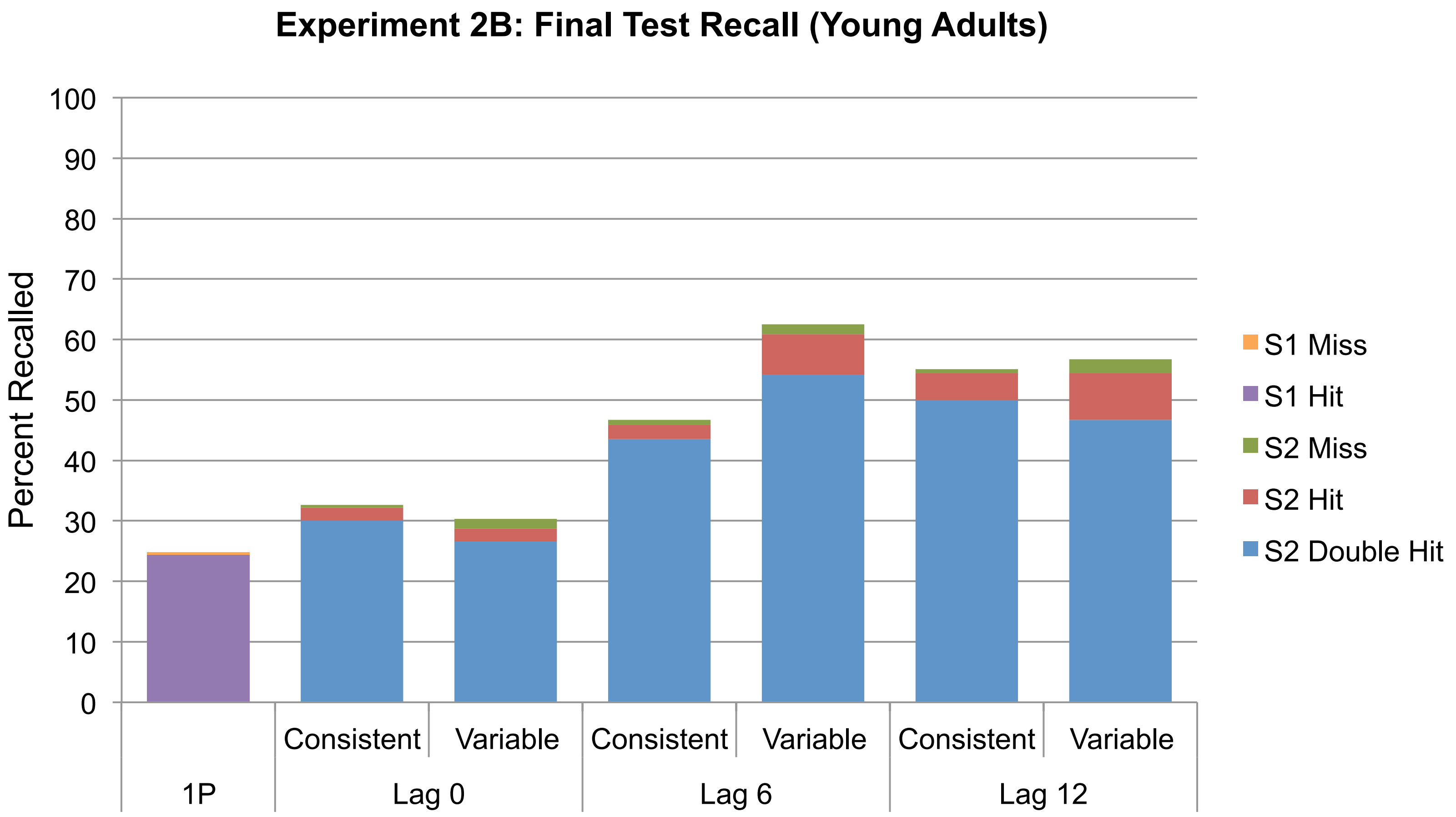

Figure 12a. Young adults' percentage of correctly recalled objects on the final test. "Hit" represents items that were correctly recognized as repeats at S2. "Miss" represents items that were not recognized as repeats at S2. "Double Hit" represents items that were correctly recognized as repeats *and* correctly recognized for their study context at S2.

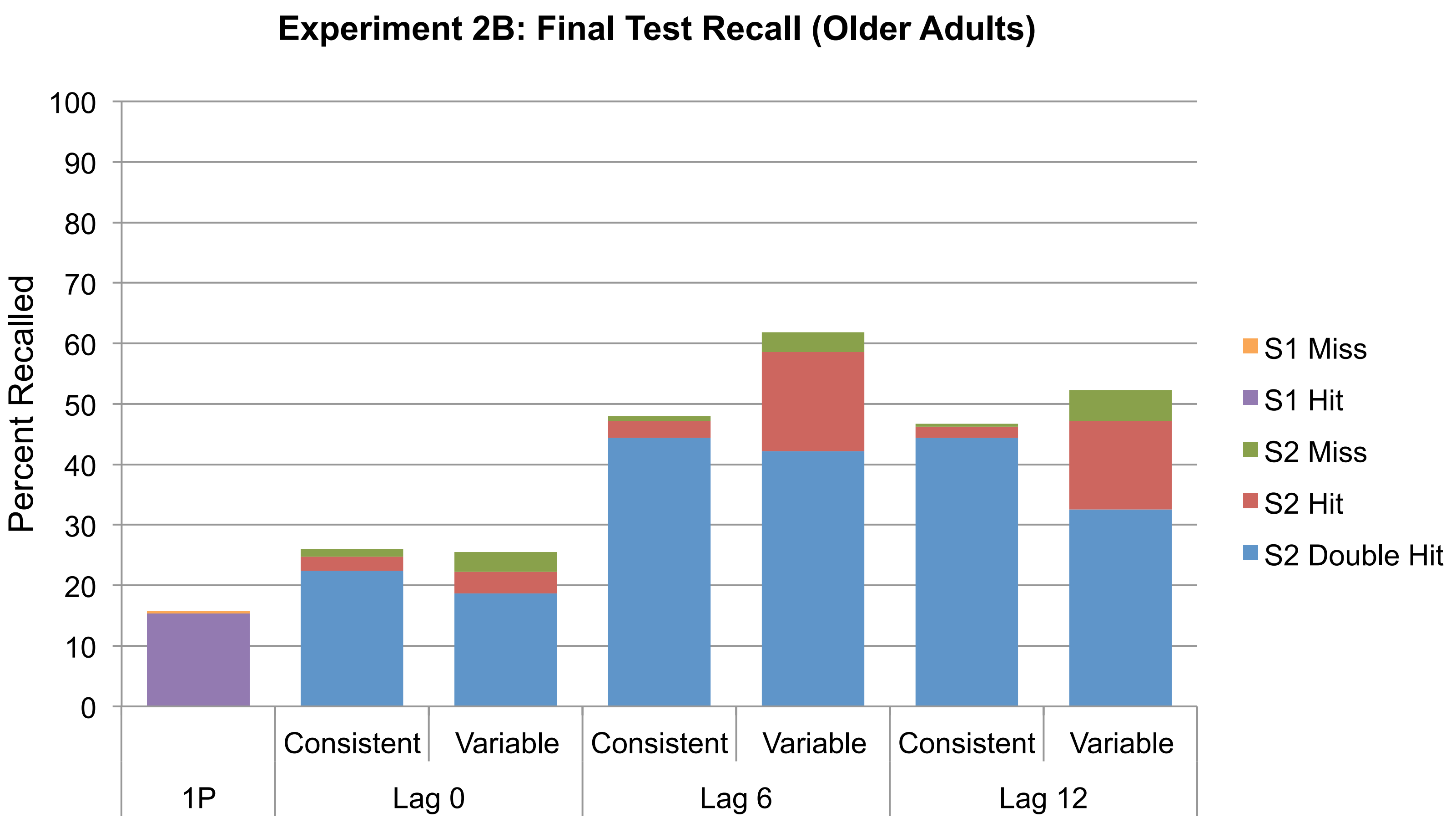

Figure 12b. Older adults' percentage of correctly recalled objects on the final test. "Hit" represents items that were correctly recognized as repeats at S2. "Miss" represents items that were not recognized as repeats at S2. "Double Hit" represents items that were correctly recognized as repeats *and* correctly recognized for their study context at S2.

62.20, *SD* = 17.24) was higher than recall of Lag 6, consistently studied items (*M* = 47.34, *SD* = 14.47), $t(68) = 7.70$, $p < .001$, $d = 0.94$, $\text{BF}_{10} = 1.24 \times 10^8$ (extreme evidence for H_1). Note that the Age × Lag × Study Context interaction was not significant.

Discussion

The main goal of Experiment 2B was to re-evaluate the Age × Lag × Study Context interaction using a paradigm where each variably studied item was associated with two unique and semantically congruent study contexts rather than two shared and semantically ambiguous study contexts. Therefore, Experiment 2A and Experiment 2B differed in their operational definition of variable study conditions. Once again, the Age × Lag × Study Context interaction in object free recall was not significant. Although there was a significant Lag × Study Context interaction, it was in the opposite direction from previous literature. These main findings are discussed below.

Age Differences in Item Memory vs. Study Context Memory

Age differences in item memory versus study context memory were less clear in Experiment 2B compared to Experiment 2A; older adults demonstrated both poorer item recognition and study context recognition compared to younger adults (93% vs. 97%, respectively and 82% vs. 90%, respectively). Age differences in item memory may have emerged in the present experiment as a result of a greater number of contextual cues presented throughout the study phase. Research has found that older adults are more likely than younger adults to “over-encode” information in their environments, including distracting and/or irrelevant information that may hinder their performance. This deficit has been attributed to failing attentional control abilities in aging (e.g., Hasher, Zacks, & May, 1999). For example, Rowe and colleagues (2006) found a priming effect in older adults (but not younger adults) for words that had been previously presented as distractors alongside pictures during an *n*-back picture task. In the present experiment, it is possible that older adults over-encoded background scene information, which, in turn, caused them greater interference during item recognition trials compared to younger adults (e.g., “Have a I seen the *basketball* before or am I confusing it with something from the previous *sports arena scene*?”) It is hypothesized that over-encoding was less likely to impact older adults’ item recognition in Experiment 2A because only two background scenes were encoded.

Similar to Experiment 2A, older adults demonstrated a similar pattern of results to younger adults of greater item/study context recognition failures for variably studied targets presented at the two distributed lags (Tables 9 and 10). As rationalized in the previous experiment, this pattern of results is somewhat challenging for the associative binding hypothesis (Naveh-Benjamin, 2000). Overall, Hypothesis 1 was not supported.

The Effects of Lag and Study Context During the Study Phase

The study phase data from Experiment 2B share similarities and differences with those reported in Experiment 2A. In terms of item recognition at S2, accuracy scores remained at ceiling. Participants were relatively impaired at recognizing items that reappeared in a different, unique background scene at both Lag 6 and Lag 12. In Experiment 2A, recognition impairments were seen only for Lag 12, variably studied items but here both distributed lags were impacted by a change in study context. Again, this finding is interpreted as a greater chance of unsuccessful study-phase retrieval for distributed, variably studied items; however, according to Bayesian statistics, it was not a robust finding.

With respect to younger adults' study context recognition at S2, the difference between Lag 0, consistently studied items and Lag 0, variably studied items was considerably smaller in this experiment (MD_{Exp2A} = 19.88 vs. MD_{Exp2B} = 3.17; Table 6 vs. Table 10). This comparison provides some evidence that a greater number of unique contextual cues improved younger participants' processing of massed, variably studied items[14]. Younger adults in this experiment also exhibited larger study context recognition impairments for variably studied items presented at the two distributed lags compared to consistently studied items presented at the two distributed lags. This finding is interpreted as further evidence of study-phase retrieval difficulty for distributed, variably studied items. With respect to older adults' study context recognition at S2, performance was once again influenced by the interaction of lag and study context, which was not a predicted pattern of results. In fact, in this experiment, variable study conditions had a relatively larger detrimental effect on older adults' S2 recognition performance compared to younger adults. Overall, these data provide mixed support for Hypothesis 2[15].

The Effects of Lag and Study Context at Final Test

The final test data documented a distributed practice effect for all participants. Items repeated at either of the two distributed lags were better recalled than items repeated at the massed lag. Unlike Experiment 2A, there was a significant Lag × Study Context interaction in this experiment, although, it was in a different direction than other

---

[14] It should be noted that the difference between older adults' study context recognition of Lag 0, consistently studied targets and Lag 0, variably studied targets was more similar between experiments (MD_{Exp2A} = 16.18 vs. MD_{Exp2B} = 11.57). This pattern of results may imply that a greater number of unique contextual cues did not assist massed learning for older adults in the same way as it did for younger adults.

[15] The additional prediction that younger adults may benefit from variable study at a distributed lag so long as the lag is optimally timed (in this case, Lag 6 but not Lag 12) was not supported by the Experiment 2B data either.

reports. Specifically, previous studies have described improvements in younger adults' subsequent memory for massed items if they are repeated under variable study conditions at S2. The interaction effect is sometimes amplified when variable study conditions have a reverse effect on subsequent memory for distributed items. Although younger adults responded favourably to massed, variably studied items during the study phase of this experiment, the same benefit was not reflected in their subsequent final recall performance. Variable study did not influence recall of Lag 0 items; rather, variable study improved recall of distributed items that had been repeated at the moderate lag (Lag 6) but not the longest lag (Lag 12). Variable study conditions had neither a positive nor a negative influence on Lag 12 items. These findings might suggest that the balance of encoding variability and study phase retrieval difficulty was optimal for Lag 6, variably studied items at S2 ($M_{\text{younger}} = 88.13$ and $M_{\text{older}} = 68.02$) relative to Lag 12, variably studied items at S2 ($M_{\text{younger}} = 81.93$ and $M_{\text{older}} = 66.86$), causing superior recall of the former on the final test.

Finally, like Experiment 2A, the Age × Lag × Study Context interaction was not significant, indicating that the effect of study context on the distributed practice effect did not operate differently between age groups. This finding is not surprising considering similar patterns of responding between age groups during the study phase. Hypothesis 3 was not supported.

General Discussion

In two experiments, the effect of study context on the distributed practice effect was examined and compared in samples of younger and older adults. Participants' performance was assessed throughout the study phase and in a final recall memory test.

In summary, a number of novel findings have been presented. First, the influence of repetition lag on subsequent memory was robust in both experiments. Lag accounted for more than half of the variance in participants' final free recall scores. The distributed practice effect is commonly explained using a hybrid account, specifically the interplay of deficient processing, encoding variability, and study-phase retrieval mechanisms. The hybrid account predicts that an item repeated in a spaced schedule is associated with a greater number of unique contextual cues simply as a result of the passage of time and that, if the item undergoes effortful (non-redundant) but successful retrieval and memory trace updating at S2, its chance of subsequent retrieval is further improved. In support of the hybrid account, both experiments found superior final recall for items repeated at the two distributed lags compared to items repeated at the massed lag. Distributed items reaped the benefits of the longest passage of time and, as indicated by the S2 data, the most effortful study-phase retrieval.

Second, the Lag × Study Context interaction previously reported for younger adults across a range of distributed practice paradigms was not replicated in either of the two experiments reported here. In Experiment 2A, there was no clear benefit to restudying a massed item in a different study context from its original presentation. In Experiment 2B, restudying a massed item in an entirely novel study context was associated with improved response accuracy during the study phase for younger adults; however, younger adults' final recall performance did not distinguish these items from massed, consistently studied items. Instead, the Lag × Study Context interaction reported in Experiment 2B reflected younger adults' superior recall of moderately distributed items (Lag 6) restudied under variable study conditions. Complementing this interaction was the finding that younger participants experienced the greatest study-phase retrieval difficulty at S2 for distributed, variably studied items.

The results of Experiment 2B imply that variable study may actually enhance the distributed practice effect under certain learning conditions. One such condition is the number of unique contextual cues present throughout the study phase. Final recall of distributed items in Experiment 2A did not depend on their study context throughout the study phase, which was limited to a set of only two background scenes. Conversely, final recall of distributed items in Experiment 2B did depend on their study context, which was characterized by several unique background scenes that could be more easily associated with each to-be-learned item in the list. The advantage of many unique contextual cues during learning on subsequent final memory is supported by Glenberg's (1979) components-level theory, which posits that to-be-remembered items are associated with contextual, structural, and descriptive components and that the most specific components (i.e., descriptive) will be most useful at the time of final retrieval. Thus, an item characterized by a greater number of unique contextual components (i.e., the effect of lag) and a greater number of unique descriptive components (i.e., the effect of unique study contexts) is most likely to be successfully retrieved on a final memory test. If learning involves only one or two study contexts, as in Experiment 2A, this information is more likely to be stored as a structural component (i.e., the structure that a learner imposes during study), which is less specific, and therefore less helpful in a future retrieval scenario.

A second condition that may influence the benefit of variable study on the distributed practice effect is the type of material being learned. The stimuli used in the present experiments were objects, which was a departure from a literature characterized primarily by verbal learning. Objects are inherently associated with richer contextual information than words (e.g., colour, shape, location, sensory information, and/or personal relevancy). It may be easier for a learner to associate objects with their contextual cues, in turn, creating more elaborate memory traces for objects compared to words. To test this hypothesis, two additional experiments could be conducted using the same experimental paradigms as Experiment 2A and 2B but substituting the object images with their labels instead. A version of this design was tested by Hockley, Bancroft, and Bryant (2012, Exp 3); however, there were noticeable floor effects that clouded interpretation of the results. Thus, until further research is conducted, it remains possible that variable study amplifies distributed practice benefits for non-verbal learning compared to verbal learning.

To summarize a third and final finding from the two experiments, a hypothesized Age × Lag × Study Context interaction was not supported by the data from either experiment. Older adults did have greater difficulty retrieving study context information at S2 compared to younger adults, especially in Experiment 2A when the same two background scenes were associated with many items. They made a greater number of errors, especially at distributed lags. However, contrary to theory, the data also suggested that older adults did bind items with their study contexts, a process which may have assisted older adults' final retrieval of the object stimuli in a similar way to younger adults. Indeed, final recall performance was comparable between age groups, as were the effects of lag and study context, and the Lag × Study Context interaction.

Conclusion

Research continues to uncover defining features of the distributed practice effect, including in what specific situations spaced repetition is most likely to bolster subsequent memory relative to massed repetition. The studies that made up this dissertation explored study context as a modifier of distributed practice benefits in an assortment of learning scenarios. Of primary interest was whether the Lag × Study Context interaction has a neural representation (Study 1) and whether the finding generalizes to an older adult population (Study 2).

Surprisingly, across three different experimental paradigms, study context did not influence the effect of lag as predicted by the hybrid account; rather, lag had a clear independent effect in all three experiments regardless of the study context in which a to-be-learned item was presented at S1/S2. In Study 1, the null interaction may have been the consequence of a recognition paradigm. As explained in the discussion of this study (p. 62), Glenberg (1979) has proposed that study context is more likely to exert influence on lag when the final test is in the form of free recall. In Study 2, the null interaction may have been the consequence of a non-verbal paradigm. As explained in the general discussion of this study (p. 114), less research has focused on the Lag × Study Context interaction in object memory, where objects are naturally associated with richer contextual cues than words. Thus, whether study context influences the distributed practice effect seems to depend on the format of the final test and the type of stimuli being learned. Importantly, hybrid account assumptions of the Lag × Study Context interaction were not refuted by the data presented across the three experiments. In particular, study phase data collected in Experiments 2A and 2B found support for the interpretation that variable study improves reprocessing of massed items at S2 while at the same time demands greater retrieval effort of distributed items at S2.

Study 1 replicated some of the findings from previous ERP studies investigating neural mechanisms of the distributed practice effect, namely that distributed repetitions may be associated with a greater recollection response at S2 (as indexed by the LPC component) relative to massed repetitions. At final test, neural responses to distributed, variably studied targets were distinguished from neural responses to non-targets (as indexed by the FN400 component); however, the former was not further distinguished from other target conditions. This latter finding is difficult to interpret in light of behavioural evidence showing that distributed, variable study was associated with superior final recognition compared to the other repetition conditions. Importantly, lag and study context did not interact to influence any of the three ERP components of interest at S2 or at final test, findings that converged with the behavioural accuracy data. Thus, the data suggested that study context imposed little/no influence on the neural representation of the distributed practice effect. As explained in the discussion of this study (p. 58), different outcomes might be observed in a follow-up experiment using a larger sample size and/or a different verbal learning paradigm.

Study 2 found that older adults behaved similarly to younger adults when reprocessing massed and distributed repetitions at S2 in either consistent or variable background scenes. Furthermore, these similar study patterns led to similar final recall performance between the two age groups. Thus, the data did not support the predicted Age × Lag × Study Context interaction. Older adults had less explicit study context knowledge than younger adults, a finding that converges with literature on aging impairments in study context memory (Spencer & Raz, 1995; Old & Naveh-Benjamin, 2008). Yet, contrary to hypotheses rooted in Naveh-Benjamin's (2000) associative deficit hypothesis, older adults did seem to engage in some degree of item and study context binding that influenced their S2 recognition performance in a manner similar to younger adults (i.e., greater study-phase retrieval difficulty for distributed, variably studied items). Thus, it seems that age, and certain age-related memory impairments, may not significantly impact the distributed practice effect.

One of the more interesting findings presented in the dissertation comes from a comparison of the results reported in Experiments 2A and 2B. These two experiments differed in their operational definition of variable study conditions, with Experiment 2A's paradigm characterized by objects presented alongside a limited number of study context cues (i.e., two background scenes) and Experiment 2B's paradigm characterized by objects presented alongside many study context cues (i.e., many background scenes). In every other respect, the two paradigms were identical. Unexpectedly, variable study *enhanced* the distributed practice effect reported in Experiment 2B, unveiled as a statistically significant Lag × Study Context interaction (extreme evidence in favour of the alternative hypothesis, according to Bayesian statistics) and with post-hoc analyses confirming that the specific benefit was for moderately spaced items. Thus, variations to study context might actually enhance the distributed practice effect in learning situations where study context cues are varied and plentiful. This idea is discussed in Glenberg's component levels theory (1979) with specific reference to subsequent free recall memory. Relatedly, the enhancement effect may further depend on the nature of the items being learned and the ease with which the learner can associate each item to its various study contexts throughout the study phase. As suggested in the general discussion of Study 2 (p. 115), a first step at untangling these questions would be to conduct two follow-up experiments using the same experimental paradigms as Experiments 2A and 2B but replacing object images with their labels instead. These data would inform whether the enhancement effect is unique to non-verbal learning or whether it generalizes to verbal learning scenarios. As a second step, it would be interesting to test a greater number of lags in the same paradigm to further explore whether variable study is optimal at a particular distributed lag (i.e., striking the right balance of encoding variability and effortful/successful study phase retrieval at S2). The present experiments attempted to explore this question but the measurement technique (i.e., recognition accuracy at S2) was not particularly sensitive. Including another measure, (e.g., recognition RTs at S2) may provide greater insight into this question.

In conclusion, the effect of distributed practice on subsequent memory continues to be one of the most investigated and yet one of the most elusive phenomena in human memory research. Whether you are a clinician, educator, or advertiser, or you are just trying to improve your golf game, understanding when and why distributed practice improves memory has widespread implications for a range of real-world learning environments.

References

Appleton-Knapp, S. L., Bjork, R. A., & Wickens, T. D. (2005). Examining the spacing effect in advertising: Encoding variability, retrieval processes, and their interaction. *Journal of Consumer Research, 32*, 266-276. https://doi.org/10.1086/432236

Balota, D. A., Dolan, P. O., & Duchek, J. M. (2000). Memory changes in healthy young and older adults. In E. Tulving & F. I. M. Craik (Eds.), *Oxford handbook of memory* (pp. 395-410). New York, NY: Oxford University Press.

Balota, D. A., Duchek, J. M., & Paullin, R. (1989). Age-related differences in the impact of spacing, lag, and retention interval. *Psychology and Aging, 4*, 3-9. https://doi.org/10.1037/0882-7974.4.1.3

Belleza, F. S., & Young, D. R. (1989). Chunking of repeated events in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 990-997. https://doi.org/10.1037/0278-7393.15.5.990

Benjamin, A. S., & Tullis, J. (2010). What makes distributed practice effective? *Cognitive Psychology, 61*, 228-247. https://doi.org/10.1016/j.cogpsych.2010.05.004

Bercovitz, K. E., Bell, M. C., Simone, P. M., & Wiseheart, M. (2017). The spacing effect in younger and older adults: Does context matter? *Aging, Neuropsychology, and Cognition, 24*, 703-716. https://doi.org/10.1080/13825585.2016.1251552

Bird, C. P., Nicholson, A. J., & Ringer, S. (1978). Resistance of the spacing effect to variations in encoding. *The American Journal of Psychology, 91*, 713-721. https://doi.org/10.2307/1421519

Bloom, K. C., & Shuell, T. J. (1981). Effects of massed and distributed practice on the learning and retention of second-language vocabulary. *Journal of Educational Research, 74,* 245-248. https://doi.org/10.1080/00220671.1981.10885317

Bower, G. H. (1972). Stimulus-sampling theory of encoding variability. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory*. Washington, DC: V. H. Winston & Sons.

Cansino, S., Estrada-Manilla, C., Hernandez-Ramos, E., Martinez-Galindo, J. G., Torres-Trejo, F., Gomez-Fernandez, T., ... Rodriguez-Oritz, M. D. (2013). The rate of source memory decline across the adult life span. *Developmental Psychology, 49*, 973-985. https://doi.org/10.1037/a0028894

Carpenter, S. K., & DeLosh, E. L. (2005). Application of the testing and spacing effects to name learning. *Applied Cognitive Psychology, 19*, 619-636. https://doi.org/10.1002/acp.1101

Carson, N., Leach, L., & Murphy, K. J. (2018). A re-examination of Montreal Cognitive Assessment (MoCA) cutoff scores. *International Journal of Geriatric Psychiatry, 33*, 379-388. https://doi.org/10.1002/gps.4756

Cepeda, N. J., Coburn, N., Rohrer, D., Wixted, J. T., Mozer, M. C., & Pashler, H. (2009). Optimizing distributed practice: Theoretical analysis and practical implications. *Experimental Psychology*, *56*, 236-246. https://doi.org/10.1027/1618-3169.56.4.236

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, *132*, 354-380. https://doi.org/10.1037/0033-2909.132.3.354

Cepeda, N. J., Vul, E., Rohrer, D., Wixted, J. T., & Pashler, H. (2008). Spacing effects in learning: A temporal ridgeline of optimal retention. *Psychological Science*, *19*, 1095-1102. https://doi.org/10.1037/0033-2909.132.3.354

Chalfonte, B. L., & Johnson, M. K. (1996). Feature memory and binding in young and older adults. *Memory & Cognition, 24*, 403-416. https://doi.org/10.3758/BF03200930

Challis, B. H. (1993). Spacing effects on cued-memory tests depend on level of processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 389-396. https://doi.org/10.1037/0278-7393.19.2.389

Chee, M. W. L., Goh, J. O. S., Venkatraman, V., Chow Tan, J., Gutchess, A., Sutton, B., ... Park, D. (2006). Age-related changes in object processing and contextual binding revealed using fMR adaptation. *Journal of Cognitive Neuroscience, 184*, 495-507. https://doi.org/10.1162/jocn.2006.18.4.495

Cherry, K. E., & Simmons-D'Gerolamo, S. S. (2005). Long-term effectiveness of spaced-retrieval memory training for older adults with probable Alzheimer's disease. *Experimental Aging Research, 31*, 261-289. https://doi.org/10.1080/0361 0730590948186

Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 33*, 497-505. https://doi.org/10.1080/14640748108400805

Crowder, R. G. (1976). *Principles of learning and memory*. Hillsdale, NJ: Erlbaum.

Criss, A. H., Aue, W., & Kilic, A. (2014). Age and response bias: Evidence from the strength-based mirror effect. *The Quarterly Journal of Experimental Psychology, 67*, 1910-1924. https://doi.org/10.1080/17470218.2013.874037

Cull, W. L., Shaughnessy, J. J., & Zechmeister, E. B. (1996). Expanding understanding of the expanding-pattern-of-retrieval mneumonic: Toward confidence in applicability. *Journal of Experimental Psychology: Applied, 2*, 365-378. https://doi.org/10.1037/1076-898X.2.4.365

Curran, T. (1999). The electrophysiology of incidental and intentional retrieval: ERP old/new effects in lexical decision and recognition memory. *Neuropsychologia, 37*, 771-785. https://doi.org/10.1016/S0028-3932(98)00133-X

Curran, T. (2000). Brain potentials of recollection and familiarity. *Memory & Cognition, 28*, 923-938. https://doi.org/10.3758/BF03209340

Curran, T., Tepe, K. L., & Piatt, C. (2006). Event-related potential explorations of dual processes in recognition memory. In H. D. Zimmer, A. Mecklinger, & U. Lindenberger (Eds.), *Binding in human memory: A neurocognitive approach* (pp. 467-492). Oxford, UK: Oxford University Press.

Dail, T. K., & Christina, R. W. (2004). Distribution of practice and metacognition in learning and long-term retention of a discrete motor task. *Research Quarterly for Exercise and Sport*, *75*, 148-155. https://doi.org/10.1080/02701367.2004.10609146

D’Agostino, P. R. (1974). Repetition and recall in an incidental-memory paradigm. *Canadian Journal of Psychology, 28*, 468-473. https://doi.org/10.1037/h0082011

D’Agostino, P. R., & DeRemer, P. (1973). Repetition effects as a function of rehearsal and encoding variability. *Journal of Verbal Learning and Verbal Behaviour, 12*, 108-113. https://doi.org/10.1016/S0022-5371(73)80066-0

Delaney, P. F., Verkoeijen, P. P. J. L., & Spirgel, A. (2010). Spacing and testing effects: A deeply critical, lengthy, and at times discursive review of the literature. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory: Vol. 53* (pp. 63-147). New York, NY: Academic Press.

Dellarosa, D. & Bourne, L. E. Jr. (1985). Surface form and the spacing effect. *Memory & Cognition, 13*, 529-537. https://doi.org/10.3758/BF03198324

Dempster, F. N. (1996). Distributing and managing the conditions of encoding and practice. In E. L. Bjork & R. A. Bjork (Eds.), *Memory: Handbook of perception and cognition* (2nd ed., pp. 317-344). San Diego, CA: Academic Press.

Denney, N. W., Miller, B. V., Dew, J. R., & Levav, A. L. (1991). An adult developmental study of contextual memory. *Journal of Gerontology: Psychological Sciences, 46,* P44-P50. https://doi.org/10.1093/geronj/46.2.P44

Dunn, J. C. (2004). Remember-know: A matter of confidence. *Psychological Review, 111*, 524-542. https://doi.org/10.1037/0033-295X.111.2.524

Durgunoglu, A. Y., Mir, M., & Arino-Marti, S. (1993). Effects of repeated readings on bilingual and monolingual memory for text. *Contemporary Educational Psychology, 18*, 294-317. https://doi.org/10.1006/ceps.1993.1022

Ebbinghaus, H. (1885/1913). *Memory: A contribution to experimental psychology* (H. A. Ruger, C. E. Bussenius, & E. R. Hilgard, Trans.). New York, NY: Dover Publications. Retrieved from http://www.psychclassics.yorku.ca

Estes, W. K. (1955). Statistical theory of distributional phenomena in learning.

Psychological Review, 62, 369-377. https://doi.org/10.1037/h0046888

Field, A. (2009). *Discovering statistics using SPSS (3rd ed).* London, UK: Sage Publications Ltd.

Friedman, D. (1990). ERPs during continuous recognition memory for words. *Biological Psychology, 30*, 61-87. https://doi.org/10.1016/0301-0511(90)90091-A

Friedman, D., & Johnson, R. Jr. (2000). Event-related potential (ERP) studies of memory encoding and retrieval: A selective review. *Microscopy Research and Technique, 51*, 6-28. https://doi.org/10.1002/1097-0029(20001001)51:1<6::AID-JEMT2>3.0.CO;2-R

Gartman, L. M., & Johnson, N. F. (1972). Massed versus distributed repetition of homographs: A test of the differential encoding hypothesis. *Journal of Verbal Learning and Verbal Behaviour, 11*, 801-808. https://doi.org/10.1016/S0022-5371(72)80016-1

Glanzer, M., & Duarte, A. (1971). Repetition between and within languages in free recall. *Journal of Verbal Learning and Verbal Behaviour, 10*, 625-630. https://doi.org/10.1016/S0022-5371(71)80069-5

Glenberg, A. M. (1976). Monotonic and nonmonotonic lag effects in paired-associate and recognition memory paradigms. *Journal of Verbal Learning and Verbal Behavior, 15*, 1-16. https://doi.org/10.1016/S0022-5371(76)90002-5

Glenberg, A. M. (1979). Components-level theory of the effects of spacing of repetitions on recall and recognition. *Memory & Cognition, 7*, 95-112. https://doi.org/10.3758/BF03197590

Glenberg, A. M., & Lehmann, T. S. (1980). Spacing repetitions over 1 week. *Memory &*

Cognition, 8, 528-538. https://doi.org/10.3758/BF03213772

Glenberg, A. M., & Smith, S. M. (1981). Spacing repetitions and solving problems are not the same. *Journal of Verbal Learning and Verbal Behavior, 20*, 110-119. https://doi.org/10.1016/S0022-5371(81)90345-5

Glover, J. A., & Corkhill, A. J. (1987). Influence of paraphrased repetitions on the spacing effect. *Journal of Educational Psychology, 79*, 198-199. https://doi.org/ 10.1037/0022-0663.79.2.198

Goh, J. O. S., Chun Siong, S., Park, D., Gutchess, A., Hebrank, A., & Chee, M. W. L. (2004). Cortical areas involved in object, background, and object-background processing revealed with functional magnetic resonance adaptation. *The Journal of Neuroscience, 24*, 10223-10228. https://doi.org/10.1523/JNEUROSCI.3373-04.2004

Grady, C. L., & Craik, F. I. M. (2000). Changes in memory processing with age. *Current Opinion in Neurobiology, 10*, 224-231. https://doi.org/10.1016/S0959-4388 (00)00073-8

Green, J. L., Weston, T., Wiseheart, M., & Rosenbaum, R. S. (2014). Long-term spacing effect benefits in developmental amnesia: Case experiments in rehabilitation. *Neuropsychology, 28*, 685-694. https://doi.org/10.1037/ neu0000070

Greene, R. L. (1989). Spacing effects in memory: Evidence for a two-process account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 371. https://doi.org/10.1037/0278-7393.15.3.371

Greene, R. L., & Stillwell, A. M. (1995). Effects of encoding variability and spacing on frequency discrimination. *Journal of Memory and Language, 34,* 468-476. https://doi.org/10.1006/jmla.1995.1021

Hasher, L., Zacks, R. T., & May, C. P. (1999). Inhibitory control, circadian arousal, and age. In D. Gopher & A. Koriat (Eds.), *Attention and performance XVII. Cognitive regulation of performance: Interaction of theory and application* (pp. 653-675). Cambridge, MA: The MIT Press.

Hintzman, D. L. (1974). Theoretical implications of the spacing effect. In R. L. Solso (Ed.), *Theories in cognitive psychology: The Loyola symposium* (pp. 77-99). Hillsdale, NJ: Erlbaum.

Hintzman, D. L. (2004). Judgment of frequency versus recognition confidence: Repetition and recursive reminding. *Memory & Cognition, 32*, 336-350. https://doi.org/10.3758/BF03196863

Hintzman, D. L., Block, R. A., & Summers, J. J. (1973). Modality tags and memory for repetitions: Locus of the spacing effect. *Journal of Verbal Learning and Verbal Behaviour, 12*, 229-238. https://doi.org/10.1016/S0022-5371(73)80013-1

Hintzman, D. L., Summers, J. J., & Block, R. A. (1975). Spacing judgment as an index of study-phase retrieval. *Journal of Experimental Psychology: Human Learning and Memory, 104*, 31-40. https://doi.org/10.1037/0278-7393.1.1.31

Hockley, W. E., Bancroft, T. D., & Bryant, E. (2012). Associative and familiarity-based effects of environmental context on memory. *Canadian Journal of Experimental Psychology, 66*, 81-89. https://doi.org/10.1037/a0027136

Huh, T. J., Kramer, J. H., Gazzaley, A., & Delis, D. C. (2006). Response bias and aging on a recognition memory task. *Journal of the International Neuropsychological Society, 12*, 1-7. https://doi.org/10.1017/S1355617706060024

Janiszewski, C., Noel, H., & Sawyer, A. G. (2003). A meta-analysis of the spacing effect in verbal learning: Implications for research on advertising repetition and consumer memory. *Journal of Consumer Research, 30*, 138-149. https://doi.org/10.1086/374692

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

Jensen, T. D., & Freund, J. S. (1981). Persistence of the spacing effect in incidental free recall: The effect of external list comparisons and inter-task correlations. *Bulletin of the Psychonomic Society, 18*, 183-186. https://doi.org/10.3758/BF03333598

Johnson, W. A., & Uhl, C. N. (1976). The contributions of encoding effort and variability to the spacing effect on free recall. *Journal of Experimental Psychology: Human Learning and Memory, 2*, 153-160. https://doi.org/10.1037/0278-7393.2.2.153

Kapler, I. V., Weston, T., & Wiseheart, M. (2015). Spacing in a simulated undergraduate classroom: Long-term benefits for factual and higher-level learning. *Learning and Instruction, 36*, 38-45. https://doi.org/10.1016/j.learninstruc.2014.11.001

Karpicke, J. D., Lehman, M., & Aue, W. R. (2014). Retrieval-based learning: An episodic context account. *Psychology of Learning and Motivation, 61*, 237-284. https://doi.org/10.1016/B978-0-12-800283-4.00007-1

Kausler, D. H., & Puckett, J. M. (1980). Adult age differences in recognition memory for a nonsemantic attribute. *Experimental Aging Research, 6*, 349-355. https://doi.org

/10.1080/03610738008258370

Kilic, A., Hoyer, W. J., & Howard, M. W. (2013). Effects of spacing of item repetitions in continuous recognition memory: Does item retrieval difficulty promote item retention in older adults? *Experimental Aging Research, 39*, 322-341. https://doi.org/10.1080/0361073X.2013.779200

Kim, M-S., Kim, J-J., & Kwon, J. S. (2001). The effect of immediate and delayed word repetition on event-related potential in a continuous recognition task. *Cognitive Brain Research, 11*, 387-396. https://doi.org/10.1016/S0926-6410(01)00011-8

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction"? *Psychological Science, 19*, 585-592. https://doi.org/ 10.1111/j.1467-9280.2008.02127.x

Küpper-Tetzel, C. E. (2014). Strong effects on weak theoretical grounds: Understanding the distributed practice effect*. Zeitschrift für Psychologie, 222*, 71-81. https://doi.org/10.1027/2151-2604/a000168

Küpper-Tetzel, C. E., Kapler, I. V., & Wiseheart, M. (2014). Contracting, equal, and expanding learning schedules: The optimal distribution of learning sessions depends on retention interval. *Memory & Cognition, 42*, 729-741. https://doi.org/ 10.3758/s13421-014-0394-1

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of event-related brain potential (ERP). *Annual Review of Psychology, 62*, 621-647. https://doi.org/10.1146/annurev.psych.093008.131123

Landauer, T. K. (1969). Reinforcement as consolidation. *Psychological Review, 76*, 82-96. https://doi.org/10.1037/h0026746

Landauer, T. K., & Bjork, R. A. (1978). Optimal rehearsal patterns and name learning. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 625-632). London, UK: Academic Press.

Logan, J. M., & Balota, D. A. (2008). Expanded vs. equal interval spaced retrieval practice: Exploring different schedules of spacing and retention interval in younger and older adults. *Aging, Neuropsychology, and Cognition, 15*, 257-280. https://doi.org/10.1080/13825580701322171

Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge, MA: MIT Press.

Luck, S. J., Woodman, G. F., & Vogel, E. K. (2000). Event-related potential studies of attention. *Trends in Cognitive Sciences, 4*, 432-440. https://doi.org/10.1016/S1364-6613(00)01545-X

Maddox, G. B. (2016). Understanding the underlying mechanism of the spacing effect in verbal learning: A case for encoding variability and study-phase retrieval. *Journal of Cognitive Psychology, 28*, 684-706. https://doi.org/10.1080/20445911.2016.1181637

Madigan, S. A. (1969). Intraserial repetition and coding processes in free recall. *Journal of Verbal Learning and Verbal Behavior*, *8*, 828-835. https://doi.org/10.1016/S0022-5371(69)80050-2

Magliero, A. (1983). Pupil dialations following pairs of identical and related to-be-remembered words. *Memory & Cognition, 11*, 609-615. https://doi.org/10.3758/BF03198285

Mammarella, N., Avons, S., & Russo, R. (2004). A short-term perceptual priming account of spacing effects in explicit cued-memory tasks for unfamiliar stimuli. *European Journal of Cognitive Psychology, 16*, 387-402. https://doi.org/10.1080/09541440340000042

Mammarella, N., Russo, R., & Avons, S. E. (2002). Spacing effects in cued-memory tasks for unfamiliar faces and nonwords. *Memory & Cognition, 30*, 1238-1251. https://doi.org/10.3758/BF03213406

Martin, E. (1968). Stimulus meaningfulness and paired-associate transfer: An encoding variability hypothesis. *Psychological Review, 75*, 421-441. https://doi.org/10.1037/h0026301

Maskarinec, A. S., & Thompson, C. P. (1976). The within-list distributed practice effect: Tests of varied context and varied encoding hypotheses. *Memory & Cognition, 4*, 741-747. https://doi.org/10.3758/BF03213242

Melton, A. W. (1967). Repetition and retrieval from memory. *Science, 158*, 532. https://doi.org/10.1126/science.158.3800.532-b

Melton, A. W. (1970). The situation with respect to the spacing of repetitions and memory. *Journal of Verbal Learning and Verbal Behaviour, 9*, 596-606. https://doi.org/10.1016/S0022-5371(70)80107-4

Menzel, R., Manz, G., Menzel, R., & Greggers, U. (2001). Massed and spaced learning in honeybees: The role of CS, US, the intertrial interval, and the test interval. *Learning and Memory, 8*, 198-208. https://doi.org/10.1101/lm.40001

Metalis, S. A. (1985). Effects of massed versus distributed practice on acquisition of video game skill. *Perceptual and Motor Skills, 61*, 457-458. https://doi.org/10.2466/pms.1985.61.2.457

Mitchell, K. J., Johnson, M. K., Raye, C. L., & D'Esposito, M. (2000). fMRI evidence of age-related hippocampal dysfunction in feature binding in working memory. *Cognitive Brain Research, 10*, 197-206. https://doi.org/10.1016/S0926-6410(00)00029-X

Mollison, M. V. (2015). *Distributed practice and distributed representations: Investigating the spacing effect using EEG* (Unpublished doctoral dissertation). University of Colorado, Boulder, Colorado.

Moulton, C. A. E., Dubrowski, A., MacRae, H., Graham, B., Grober, E., & Reznick, R. (2006). Teaching surgical skills: What kind of practice makes perfect? *Annals of Surgery, 244*, 400-409. https://doi.org/10.1097/01.sla.0000234808.85789.6a

Mozer, M. C., Pashler, H., Cepeda, N. J., Lindsey, R., & Vul, E. (2009). Predicting the optimal spacing of study: A multiscale context model of memory. In Y. Bengio, D. Schuurmans, J. Lafferty, C. K. I. Williams, & A. Culotta (Eds.), *Advances in neural information processing systems: Vol. 22* (pp. 1321-1329). La Jolla, CA: Neural Information Processing Systems Foundation.

Nagy, M. E., & Rugg, M. D. (1989). Modulation of event-related potentials by word repetition: The effects of inter-item lag. *Psychophysiology, 26*, 431-436. https://doi.org/10.1111/j.1469-8986.1989.tb01946.x

Nasreddine, Z. S., Phillips, N. A., Bedirian, V., Charbonneau, S., Whitehead, V., Collin, I., ... Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society, 53*, 695-699. https://doi.org/10.1111/j.1532-5415.2005.53221.x

Naveh-Benjamin, M. (2000). Adult age differences in memory performance: Tests of an associative deficit hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 1170. https://doi.org/10.1037/0278-7393.26.5.1170

Naveh-Benjamin, M., & Craik, F. I. M. (1995). Memory for context and its use in item memory: Comparisons of younger and older persons. *Psychology and Aging, 10*, 284-293. https://doi.org/10.1037/0882-7974.10.2.284

Old, S. R., & Naveh-Benjamin, M. (2000). Differential effects of age on item and associative measures of memory. *Psychology and Aging, 23*, 104-118. https://doi.org/10.1037/0882-7974.23.1.104

Paivio, A. (1974). Spacing of repetitions in the incidental and intentional free recall of pictures and words. *Journal of Verbal Learning and Verbal Behavior, 13*, 497-511. https://doi.org/10.1016/S0022-5371(74)80002-2

Paivio, A., Clark, J. M., & Lambert, W. E. (1988). Bilingual dual-coding theory and semantic repetition effects on recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 163-172. https://doi.org/10.1037/0278-7393.14.1.163

Paivio, A. & Csapo, K. (1973). Picture superiority in free recall: Imagery or dual coding. *Cognitive Psychology, 5*, 176-206. https://doi.org/10.1016/0010-0285(73)90032-7

Park, D. C., & Puglisi, J. T. (1985). Older adults' memory for the color of pictures and words. *Journal of Gerontology, 40*, 198-204. https://doi.org/10.1093/geronj/40.2.198

Park, D. C., Puglisi, J. T., & Sovacool, M. (1983). Memory for pictures, words, and spatial location in older adults: Evidence for pictorial superiority. *Journal of Gerontology, 38*, 582-588. https://doi.org/10.1093/geronj/38.5.582

Peterson, L. R. (1966). Short-term verbal memory and learning. *Psychological Review, 73*, 193-207. https://doi.org/10.1037/h0023229

Peterson, L. R., Wampler, R., Kirkpatrick, M., & Saltzman, D. (1963). Effect of spacing presentations on retention of a paired associate over short intervals. *Journal of Experimental Psychology, 66*, 206-209. https://doi.org/10.1037/h0046694

Posner, M. I., & McCandliss, B. D. (1999). Brain circuitry during reading. In R. M. Klein & P. McMullen (Eds.), *Converging methods for understanding reading and dyslexia* (pp. 305-337). Cambridge, MA: MIT Press.

Pratt, H. (2012). Sensory ERP components. In S. J. Luck & E. S. Kappenman (Eds.), *The Oxford handbook of event-related component potentials* (pp. 89-114). New York, NY: Oxford University Press.

Pyc, M. A., Balota, D. A., McDermott, K. B., Tully, T., & Roediger, H. L. III. (2014). Between-list lag effects in recall depend on retention interval. *Memory & Cognition, 42*, 965-977. https://doi.org/10.3758/s13421-014-0406-1

Raaijmakers, J. G. W. (2003). Spacing and repetition effects in human memory. Application of the SAM model. *Cognitive Science, 27*, 431-452. https://doi.org/10.1207/s15516709cog2703_5

Rawson, K. A., & Kintsch, W. (2005). Rereading effects depend on time of test. *Journal of Educational Psychology, 97*, 70-80. https://doi.org/10.1037/0022-0663.97.1.70

Raz, N., Lindenberger, U., Rodrigue, K. M., Kennedy, K. M., Head, D., Williamson, A., ... & Acker, J. D. (2005). Regional brain changes in healthy aging adults: General trends, individual differences and modifiers. *Cerebral Cortex, 15*, 1676-1689. https://doi.org/10.1093/cercor/bhi044

Rohrer, D., & Taylor, K. (2007). The shuffling of mathematics problems improves learning. *Instructional Science, 35*, 481-498. https://doi.org/10.1007/s11251-007-9015-8

Rose, R. J. (1980). Encoding variability, levels of processing, and the effects of spacing of repetitions upon judgments of frequency. *Memory & Cognition, 8*, 84-93. https://doi.org/10.3758/BF03197555

Rose, R. J. (1984). Processing time for repetitions and the spacing effect. *Canadian Journal of Psychology, 83*, 537-550. https://doi.org/10.1037/h0080870

Ross, B. H., & Landauer, T. K. (1978). Memory for at least one of two items: Test and failure of several theories of spacing effects. *Journal of Verbal Learning and Verbal Behaviour, 17*, 669-676. https://doi.org/10.1016/S0022-5371(78)90403-6

Rovee-Collier, C., Evancio, S., & Earley, L. A. (1995). The time window hypothesis: Spacing effects. *Infant Behavior and Development, 18*, 69-78. https://doi.org/10.1016/0163-6383(95)90008-X

Rowe, G., Valderrama, S., Hasher, L., & Lenartowicz, A. (2006). Attentional disregulation. A benefit for implicit memory. *Psychology and Aging, 21*, 826-830. https://doi.org/10.1037/0882-7974.21.4.826

Rugg, M. D. (1995). ERP studies of memory. In M. D. Rugg & M. G. H. Coles (Eds.), *Electrophysiology of the mind: Event-related brain potentials and cognition* (pp. 132-170). Oxford, UK: Oxford University Press.

Rugg, M. D., & Allan, K. (2000). Event-related potential studies of memory. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 521-537). New York, NY: Oxford University Press.

Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. *Trends in Cognitive Sciences, 11*, 251-257. https://doi.org/10.1016/j.tics.2007.04.004

Rugg, M. D., & Doyle, M. C. (1992). Event-related potentials and recognition memory for low- and high-frequency words. *Journal of Cognitive Neuroscience, 4*, 69-79. https://doi.org/10.1162/jocn.1992.4.1.69

Rugg, M. D., Mark, R. E., Walla, P., Schloerscheidt, A. M., Birch, C. S., & Allan, K. (1998). Dissociations of the neural correlates of implicit and explicit memory. *Nature, 392*, 595-598. https://doi.org/10.1038/33396

Rundus, D. (1971). Analysis of rehearsal processes in free recall. *Journal of Experimental Psychology, 89*, 63-77. https://doi.org/10.1037/h0031185

Russo, R., Mammarella, N., & Avons, S. E. (2002). Toward a unified account of spacing effects in explicit cued-memory tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 819-829. https://doi.org/10.1037/0278-7393.28.5.819

Russo, R., Parkin, A. J., Taylor, S. R., & Wilks, J. (1998). Revising current two-process accounts of spacing effects in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 161-172. https://doi.org/10.1037/0278-7393.24.1.161

Schacter, D. L., Kaszniak, A. W., Kihlstrom, J. F., & Valdiserri, M. (1991). The relation between source memory and aging. *Psychology and Aging, 6*, 559-568. https://doi.org/10.1037/0882-7974.6.4.559

Schacter, D. L., Wig, G. S., & Stevens, W. D. (2007). Reductions in cortical activity during priming. *Current Opinion in Neurobiology, 17*, 171-176. https://doi.org/10.1016/j.conb.2007.02.001

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3*, 207-217. https://doi.org/10.1111/j.1467-9280.1992.tb00029.x

Segalowitz, S. J., Van Roon, P., & Dywan, J. (1997). The ERP late positivity: A graduated response to stimulus repetition. *NeuroReport, 8*, 757-760. https://doi.org/10.1097/00001756-199702100-00035

Shaughnessy, J. J. (1976). Persistence of the spacing effect in free recall under varying incidental learning conditions. *Memory & Cognition, 4*, 369-377. https://doi.org/10.3758/BF03213192

Shaughnessy, J. J., Zimmerman, J., & Underwood, B. J. (1972). Further evidence on the MP-DP effect in free-recall learning, *Journal of Verbal Learning and Verbal Behaviour, 11*, 1-12. https://doi.org/10.1016/S0022-5371(72)80053-7

Shea, C. H., Lai, Q., Black, C., & Park, J-H. (2000). Spacing practice sessions across days benefits the learning of motor skills. *Human Movement Science, 19*, 737-760. https://doi.org/10.1016/S0167-9457(00)00021-X

Shipley, W. C., Gruber, C. P., Martin, T. A., & Klein, A. M. (2009). *Shipley-2 manual*. Los Angeles, CA: Western Psychological Services.

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*, 755-764. https://doi.org/10.1037/a0035585

Simmons, A. L. (2012). Distributed practice and procedural memory consolidation in musicians' skill learning. *Journal of Research in Music Education, 59*, 357-368. https://doi.org/10.1177/0022429411424798

Simone, P. M., Bell, M. C., & Cepeda, N. J. (2012). Diminished but not forgotten: Effects of aging on magnitude of spacing effect benefits. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 68*, 674-680. https://doi.org/10.1093/geronb/gbs096

Smith, T., Gildeh, N., & Holmes, C. (2007). The Montreal Cognitive Assessment: Validity and utility in a memory clinic setting. *The Canadian Journal of Psychiatry, 52*, 329-332. https://doi.org/10.1177/070674370705200508

Sobel, H. S., Cepeda, N. J., & Kapler, I. V. (2011). Spacing effects in real-world classroom vocabulary learning. *Applied Cognitive Psychology, 25*, 763-767. https://doi.org/10.1002/acp.1747

Spencer, W. D., & Raz, N. (1995). Differential effects of aging on memory for content and context: A meta-analysis. *Psychology and Aging, 10*, 527-539. https://doi.org/10.1037/0882-7974.10.4.527

Thios, S. J. (1972). Memory for words in repeated sentences. *Journal of Verbal Learning and Verbal Behavior, 11*, 789-793. https://doi.org/10.1016/S0022-5371(72)80014-8

Thios, S. J., & D'Agostino, P. R. (1976). Effects of repetition as a function of study-phase retrieval. *Journal of Verbal Learning and Verbal Behavior, 15,* 529-536. https://doi.org/10.1016/0022-5371(76)90047-5

Toppino, T. C., & DeMesquita, M. (1984). Effects of spacing repetitions on children's memory. *Journal of Experimental Child Psychology, 37*, 637-648. https://doi.org/

10.1016/0022-0965(84)90081-X

Toppino, T. C., & Gerbier, E. (2014). About practice: Repetition, spacing, and abstraction. In B. H. Ross (Ed.), *Psychology of Learning and Motivation: Vol. 60* (pp. 113-189). Waltham, MA: Academic Press. https://doi.org/10.1016/B978-0-12-800090-8.00004-4

Tulving, E. (1968). When is recall higher than recognition? *Psychonomic Science, 10*, 53-54. https://doi.org/10.3758/BF03331403

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology, 26*, 1-12. https://doi.org/10.1037/h0080017

Tulving, E. & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*, 352-373. https://doi.org/10.1037/h0020071

Underwood, B. J. (1969). Some correlates of item repetition in free-recall learning. *Journal of Verbal Learning and Verbal Behaviour, 8*, 84-94. https://doi.org/10.1016/S0022-5371(69)80015-0

Van Strien, J. W., Verkoeijen, P. P. J. L., Van der Meer, N., & Franken, I. H. A. (2007). Electrophysiological correlates of word repetition spacing: ERP and induced band power old/new effects with massed and spaced repetitions. *International Journal of Psychophysiology, 66*, 205-214. https://doi.org/10.1016/j.ijpsycho.2007.07.003

Verkoeijen, P. P. J. L. (2005). *Explaining the spacing effect: Study-phase retrieval, contextual-variability and priming accounts* (Unpublished doctoral dissertation). Erasmus University Rotterdam, The Netherlands.

Verkoeijen, P. P. J. L., Rikers, R. M. J. P., & Schmidt, H. G. (2004). Detrimental influence of contextual change on spacing effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 796-800. https://doi.org/10.1037/0278-7393.30.4.796

Verkoeijen, P. P. J. L., Rikers, R. M. J. P., & Schmidt, H. G. (2005). Limitations to the spacing effect: Demonstration of an inverted u-shaped relationships between interrepetition spacing and free recall. *Experimental Psychology, 52*, 257-263. https://doi.org/10.1027/1618-3169.52.4.257

Vogel, E. K., & Luck, S. J. (2000). The visual N1 component as an index of a discrimination process. *Psychophysiology, 37*, 190-203. https://doi.org/10.1111/1469-8986.3720190

von Wright, J. M. (1976). Effects of context variation, intracategory variation, and spacing on the free recall of pictures. *Scandinavian Journal of Psychology, 17*, 303-308. https://doi.org/10.1111/j.1467-9450.1976.tb00244.x

Waggoner, T. L. (2005). *Pseudoisochromatic plates ishihara compatible (PIPIC): 24 plate edition*. Gulf Breeze, FL: T. L. Waggoner Inc.

Wechsler, D. (2008). *Wechsler adult intelligence scale--Fourth edition (WAIS-IV)*. San Antonio, TX: NCS Pearson.

Wells, J. E., & Kirsner, K. (1974). Repetition between and within modalities in free recall. *Bulletin of the Psychonomic Society, 2*, 395-397. https://doi.org/10.3758/BF03336730

Wiggs, C. L., & Martin, A. (1998). Properties and mechanisms of perceptual priming. *Current Opinion in Neurobiology, 8*, 227-233. https://doi.org/10.1016/S0959-4388(98)80144-X

Wilding, E. L., & Rugg, M. D. (1996). An event-related potential study of recognition memory with and without retrieval of source. *Brain, 119*, 889-905. https://doi.org/10.1093/brain/119.3.889

Winograd, E., & Raines, S. R. (1972). Semantic and temporal variation in recognition memory. *Journal of Verbal Learning and Verbal Behaviour, 11*, 114-119. https://doi.org/10.1016/S0022-5371(72)80067-7

Wiseheart, M., Küpper-Tetzel, C. E., Weston, T., Kim, A. S. N., Kapler, I. V., & Foot, V. (in press). Enhancing the quality of student learning using distributed practice. In J. Dunlosky & K. Rawson (Eds.), *Cambridge handbook of cognition and education*.

Xue, G., Mei, L., Chen, C., Lu, Z-L., Poldrack, R., & Dong, Q. (2011). Spaced learning enhances subsequent recognition memory by reducing neural repetition suppression. *Journal of Cognitive Neuroscience, 23*, 1624-1633. https://doi.org/10.1162/jocn.2010.21532

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46*, 441-517. https://doi.org/10.1006/jmla.2002.2864