Signed ____

Zachary Scott Kauffman……………………………………………

Date ____________________

The Roles of Encoding Variability and Reminding in the Spacing Effect

Zachary Scott Kauffman

Department of Psychology
Rhodes College
Memphis, Tennessee

2017

Submitted in partial fulfillment of the requirements for the
Bachelor of Arts degree with Honors in Psychology

This Honors paper by Zachary Scott Kauffman has been read and approved for Honors in Psychology.

Dr. Geoffrey Maddox
(Project Advisor)

Dr. Katherine White
(Department of Psychology Chair)

Dr. Jonathan Cook
(Department of Psychology)

Dr. Kelly Weeks
(Department of Business)

CONTENTS

ABSTRACT

The Roles of Encoding Variability and Reminding in the Spacing Effect

by

Zachary Scott Kauffman

The spacing effect is the robust finding that long-term memory performance is better when material has been studied under spaced versus massed (e.g., crammed) conditions. The current study investigated the extent to which two different mechanistic accounts (i.e., reminding with desirable difficulty and reminding with encoding variability) can adequately explain the spacing effect. Across two experiments, word pairs were presented two times separated by either zero, three or ten intervening pairs. Repetitions of each word pair were presented either in the same or in different contexts across presentations. The type of reminding was manipulated across experiments. In Experiment 1, participants were consciously reminded to mentally “look back” through the study list for repetitions of items by being asked if the currently presented item had been previously studied earlier in the list. In Experiment 2, reminding was automatic during encoding because participants were not asked to “look back” across the study list for repetitions of items. Results from Experiment 1 revealed that the response latency for reminding was not predictive of final test performance. Instead results were consistent with the assumptions of encoding variability in the massed and short spacing conditions, but not the long spacing condition. Results from Experiment 2 further supported the encoding variability account across all massed and spacing conditions. Together, results suggest a minimal role of desirable difficulty and a potential benefit of encoding variability that is qualified by reminding quality in terms of recollection and familiarity as a basis for reminding.

Introduction

The *spacing effect* refers to a phenomenon in which long-term memory performance is enhanced for a wide range of material (e.g., words, pictures) and a wide range of populations (e.g., children, young adults, older adults) as a result of spaced study sessions relative to massed study sessions (i.e., cramming; see Maddox, 2016 for a review). Research investigating the spacing effect has typically revealed a nonmonotonic function relating lag (i.e., the interval between repeated study events) to long-term memory performance (Cepeda, Pashler, Vul, Wixted & Rohrer, 2006). Specifically, memory performance increases as lag increases until an optimal lag is reached, wherein memory performance peaks. As the lag continues to increase beyond the optimal lag, memory performance decreases (see Figure 1). Thus, the function describing the relationship between lag and memory performance is an inverted U-shape.

Despite over 100 years of research on the spacing effect (Ebbinghaus, 1885), the mechanism underlying this benefit has not been clearly identified. Though several mechanisms have been proposed to explain the spacing effect, the current study focused on two leading accounts: Benjamin and Tullis' (2010) dual-mechanism combining desirable difficulties (Bjork, 1994) and reminding (Hintzman, 2004; 2010), and Greene's (1989) dual-mechanism combining encoding variability (Melton, 1967; 1970) and reminding. Both mechanisms have been proposed and defended using largely circular reasoning, and thus the goal of the current study was to better evaluate the extent to which these two mechanisms can contribute to the memory benefit of spacing by directly manipulating encoding variability while also operationally defining and measuring reminding difficulty. Memory could be enhanced through a better understanding of the spacing effect’s underlying mechanism in classrooms (i.e., young adults learning course material), professional workplaces (i.e., young and middle-aged adults in workplace training programs), and for individuals with cognitive impairments (e.g., healthy older adults and individuals with Alzheimer’s disease).

Given that reminding is included in both accounts, I consider this mechanism first. Then I consider the mechanisms that differ between Greene’s (1989) and Benjamin and Tullis’ (2010) accounts – encoding variability and desirable difficulty, respectively.

Reminding

The reminding mechanism suggests that processing an item can trigger the detection, or reminding, of earlier instances of the same or related items (e.g., Benjamin & Tullis, 2010), and in terms of the spacing effect, the critical assumption is that the benefit of spaced study is obtained in final test performance for repeated items that were detected as repetitions during the learning phase. Specifically, when reminding is successful, it is assumed that the memory trace established during the item’s first presentation is updated and elaborated with information available at the time of repetition. In contrast, when reminding fails, a new and independent memory trace will be established that consists only of information available at the time of second presentation (Benjamin & Tullis, 2010; see also Ross & Landauer, 1978). Thus, the reminding account assumes that successful reminding produces a more robust trace than reminding failure, and as such the reminding mechanism predicts greater retrievability of these word at final test. Indeed, failing to detect an item as a repetition often reduces or eliminates the memory benefit afforded by repeated study (Madigan, 1969) as well as the memory benefit gained from spaced study (Thios & D’Agostino, 1976).

Given the assumption that reminding is critical for obtaining the benefit of spacing in long-term memory performance, it is important to consider the learning conditions in which reminding is likely to succeed and fail. Generally it is assumed that the probability of successful reminding decreases with increases in lag. As the interval between the first and second presentation of an item increases, the potential of the second representation of the item to trigger detection of the first presentation of the item decreases, because the memory trace representing the first presentation of the item becomes increasingly susceptible to interference and forgetting. Interference refers to the inability to accurately detect the reminding target (i.e., the first presentation of an item) due to the disruption, or interference, of irrelevant information that is also stored in memory (i.e., interference theory, see Underwood & Postman, 1960). On the other hand, forgetting refers to the decay, or loss, of the memory trace representing the first presentation of an item (i.e., trace decay theory of forgetting; see Brown, 1958). With this in mind, consider the study list in Figure 2 on the following page. In this scenario, the probability of detecting the repetitions of *DICE* (Lag 3) would be greater than the probability of detecting repetitions of *PLANE* (Lag 10), because there are a greater number of intervening items in the Lag 10 condition relative to the Lag 3 condition. Of course, detection of *FRUIT* (Lag 0) as a repetition should be near perfect given that there is limited opportunity for forgetting or interference in the Lag 0 condition. In many instances it is assumed that reminding occurs spontaneously such that detection of repetitions occurs without the intention of the learner and otherwise “pops” into the learner’s head (i.e., spontaneous reminding). In other instances reminding may occur in a controlled, intentional fashion such that the learner is directed to look for repetitions (i.e., controlled reminding). One might expect that intentionally “looking back” through the list for repetitions may allow reminding to occur after longer lags than conditions in which reminding occurs spontaneously, and indeed, this expectation is supported (e.g., Jacoby, 1974; Wahlheim, Maddox, & Jacoby, 2014). As such, both lag and the type of reminding utilized will influence the probability of successful reminding.

Regardless of whether reminding occurs in controlled or spontaneous form, it is critical to note that reminding cannot account for the spacing effect when considered as an independent mechanism. As previously stated, the probability of success of a reminding event, whether occurring spontaneously or in a controlled fashion, decreases as the interval separating presentations of an item increases. In turn, one would expect a monotonically decreasing function relating spacing to final test performance. Clearly this is not the ubiquitous function displayed in Figure 1. As such, the reminding mechanism can account for the decreasing portion of the nonmonotonic function relating lag and long-term memory performance, but it cannot account for the increasing portion of this function. Consequently, reminding has been combined with encoding variability in Greene’s (1989) dual-mechanism and with desirable difficulty in Benjamin & Tullis’ (2010) dual-mechanism. Encoding variability and desirable difficulty both attempt to explain the increasing portion of the spacing effect function and will be considered in turn.

Encoding Variability

One proposed method for explaining the increasing portion of the function relating lag and memory performance is to incorporate encoding variability (e.g., Glenberg, 1976; Glenberg, 1979; Melton, 1970) with reminding to produce a dual-mechanism account of the spacing effect (e.g., Greene, 1989; Raaijmakers, 2003). The encoding variability mechanism suggests that the long-term memory benefit resulting from increased lag between study events is due to change in contexts (e.g., more unique list items occurring before and after the target item) across presentations of the same item compared to a massed condition in which the contexts across repetitions should be highly similar or even the same (e.g., the surrounding list items should be more similar across presentations of the target item). In turn, variable contexts are assumed to aid long-term memory performance, because the increase in unique contextual components during encoding is expected to produce an increase in retrieval cues at the time of final test.

There are multiple ways in which context may vary across repetitions of items. Specifically, Glenberg (1979) suggested that three components of variable encoding may be encountered: contextual components, structural components, and descriptive components. Contextual components refer to the physical components of the study environment and are assumed to influence the learner's processing of the critical stimulus automatically. Structural components arise from the controlled processing of the to-be-remembered information and include the associations formed between study items. In this sense, the influence of structural components on encoding reflects intentional, effortful processing on the learner's part to relate one stimulus with another stimulus. Finally, descriptive components arise from the controlled processing of the to-be-remembered information and include the lexical and semantic characteristics of the study material. Thus, the influence of descriptive components on encoding may reflect the learner's intentional choice to think about different (or similar) semantic meanings and item features of a given stimulus. In sum, encoding variability is expected to produce the increasing portion of the function relating lag to long-term memory performance and it does so through the encoding of components of variability as per Glenberg's (1979) multi-component theory.

Although most studies rely on the assumption that repetitions separated by longer lags are surrounded by more unique list items than repetitions that occur in massed fashion (see Balota, Duchek, & Paullin, 1989, and Maddox, Pyc, Gatewood, Kauffman, & Schonhoff, in prep, for supporting evidence), a select set of studies have intentionally manipulated variability across repetitions throughout the learning phase (Gartman & Johnson, 1972; Hintzmann, 1972). For example, Verkoeijen, Rikers, and Schmidt (2005) manipulated encoding variability on the level of contextual components using the background color on which each item was presented. For variable encoding, an item was presented twice on different background colors (e.g., presented first on a red background and presented again on a green background), whereas a constant encoded item was presented twice on the same background color (e.g., presented on a red background for the first and second presentation). Repetitions occurred following either a lag of zero intervening items (i.e., massed) or six intervening items, and results revealed an isolated benefit of encoding variability at Lag 0. In other words, the optimal lag for items studied under variable encoding was shorter than the optimal lag for items studied under constant encoding.

Critically, encoding variability cannot predict the downward turn in memory performance observed in the relationship between lag and long-term memory performance. Specifically, as lag increases, the context is assumed to become more variable which in turn creates more variable retrieval routes for later retrieval. Even at long lags, encoding variability predicts a steady or asymptotic increase in memory performance – increasing the variability in which items are encoded increases the number of retrieval routes available at retrieval, increasing the probability that retrieval will be successful. Again, the reminding mechanism predicts the decreasing portion of the function, as the lag reaches a point in which reminding is unsuccessful. Simply put, Greene's (1989) two-process account claims that long-term memory performance increases as lag increases due to increasingly variable contexts surrounding repeated study events, until which point reminding fails and memory performance decreases.

Desirable Difficulty

An alternative mechanism proposed by Benjamin and Tullis (2010) accounts for the spacing effect by combining the reminding mechanism with the desirable difficulty mechanism (Bjork, 1994). At its base level, the desirable difficulty account posits that learning events that are more difficult and require more effort will lead to better long-term memory than learning events that are relatively easy in nature. Difficulty can be introduced during the learning process in various forms, and in many instances, *desirable difficulty* is similar to a manipulation of levels of processing (e.g., Craik & Tulving, 1975). However, this is not always the case. Consider desirable difficulty with regard to the spacing effect. As items are repeated across a list of to-be-remembered items, one may assume that the learner is intentionally studying the items with the use of an encoding strategy that relies heavily on a deep level of processing, and we would expect that this is true regardless of whether items are separated by a short lag, a moderate lag, or a long lag. Instead, it is the ease with which reminding occurs that contributes to desirable difficulty. Specifically, Benjamin and Tullis (2010, pg. 239) suggest that the act of reminding "potentiates memory" such that the benefit afforded to long-term memory performance for successfully reminded items is "positively related to the difficulty of the [reminding].” They justify this claim by stating that “successful reminding following a high degree of forgetting potential…enhances memory more for the [reminded] information than would a reminding following little forgetting potential.” This forgetting potential is computed by the power-law of forgetting (Wixted & Carpenter, 2007), which describes the increase in forgetting potential with increases in the time passed since the study event. Thus, desirable difficulty as a single mechanism predicts a monotonically increasing function relating lag to final test performance or a function that approaches asymptote at longer lags. However, it cannot explain the downward turn in the ubiquitous nonmonotonic function (Figure 1).

One way to address desirable difficulty’s inability to account for the spacing effect is to combine desirable difficulty with a reminding mechanism (Benjamin & Tullis, 2010). If the effects of desirable difficulty in the context of the spacing effect are dependent on forgetting potential, then the increasing benefit of desirable difficulty should be observed as the lag between repetitions increases to the extent that forgetting is so great that reminding during the acquisition phase is unsuccessful. In turn, the item will not benefit from repeated study. Thus, the suggestion that the reminding and desirable difficulty mechanisms operate concurrently resolves the shortcomings of each mechanism individually. Nevertheless, given its circular nature, the desirable difficulty mechanism is reminiscent of the Goldilocks fairy tale in the sense that researchers are attempting to identify the “sweet spot” of spacing. However, it is important to note that it is hard to determine a priori which conditions will be “too easy,” which conditions will be “too hard,” and which conditions will be “just right.” The difficulty identifying ideal conditions a priori is reflected in Benjamin and Tullis’ (2010) inability to define “reminding difficulty” in objective terms that would allow for a clear and conclusive test of the theory in the context of the spacing effect. This concern is further observed in that Benjamin and Tullis’ (2010) meta-analysis is primarily concerned with research in the cognitive literature that also fails to operationalize desirable difficulty. Very few studies have attempted to track desirable difficulty in a measurable way. Notably, the research that has attempted to examine this mechanism in a measurable way has relied on response latencies of the reminding event. In other words, the longer it takes for participants to detect a repetition, the harder the reminding event must have been (e.g., Karpicke & Roediger, 2007; Logan & Balota, 2008; Maddox & Balota, 2015; Maddox et al., in prep; Pyc & Rawson, 2009). Thus, I turn to a discussion of recent work that utilized this approach before considering the aims of the current study.

Recent Findings

In an attempt to more directly track the contributions of desirable difficulty to the benefit of spaced study in long-term memory, Maddox et al. (in prep) compared recognition and free recall performance for material separately by two relatively similar lags during the learning phase. Importantly, researchers brought reminding under conscious control by prompting participants to judge whether an item had been presented before on each trial of the learning phase. To examine the contributions of desirable difficulty to the spacing effect, reminding difficulty was operationalized as the response time taken to successfully detect a repetition. Reminding difficulty was operationalized in this way because a more effortful reminding event should take longer than a less effortful reminding event.

Long-term memory performance was assessed with a final recognition test (Experiment 1) or a free recall test (Experiments 2 and 3) for previously studied repeated items that were separated by one or five intervening items. In determining reminding difficulty, analyses utilized only reminding attempts during the acquisition phase that were successful. Results from these experiments revealed higher final recognition test accuracy for items detected as a repetition with faster response latencies than items detected as a repetition with slower response latencies during the acquisition phase. In other words, for all items that were successfully identified as repetitions (i.e., reminding was successful), long-term memory test performance benefitted more when reminding attempts were less effortful relative to when reminding attempts were more effortful during the acquisition phase. Given that the desirable difficulty mechanism suggests that long-term memory performance greater for items with more effortful reminding relative to less effortful reminding, this pattern of results observed across recognition and recall performance fails to support the desirable difficulty mechanism. However, post hoc analyses suggested that the obtained results were consistent with the encoding variability mechanism. Specifically, post hoc analyses compared the number of unique items appearing on either side of the Lag 1 items with the number of unique items appearing on either side of the Lag 5 items, revealing a significantly greater number of unique items surrounding Lag 5 items relative to Lag 1 items. Given that novel items surrounding the critical lag items should facilitate the formation of more unique inter-item associations and more variable semantic processing of the critical stimuli, (i.e., encoding variability via structural and descriptive components), these post hoc analyses suggest a role of encoding variability in producing the spacing effect observed in this study. However, because encoding variability was not manipulated in this study, it is critical to examine the influence of Greene’s (1989) alternative account utilizing an approach in which encoding variability is explicitly manipulated *a priori*.

Current Study

The current study addressed two aims. The first aim was to investigate the mechanisms underlying the spacing effect by dissociating Benjamin and Tullis’ (2010) account combining the desirable difficulty and reminding mechanisms from Greene’s (1989) account combining the encoding variability and reminding mechanisms. The current study will be the first to my knowledge to simultaneously measure reminding response latency as a proxy for desirable difficulty while explicitly manipulating encoding context to produce variable and constant encoding conditions.

The second aim of the current study was to investigate the relationship between reminding and encoding variability. Although Greene’s (1989) two-process account predicts that long-term memory performance increases as the variability between the first and second presentations of an item increases, it is still unclear what information is incorporated into the memory representation created during the first presentation of an item when that item is reminded. Some research suggests that temporal ordering (i.e., remembering which item of two related items was presented first; Hintzman, 2004) and the number of presentations of an item are entered and preserved in the memory trace more so automatically rather than controlled. One may also expect, per previous research (e.g., Glenberg, 1979), that components of variability will be combined across repetitions of an item in its underlying memory representation. Thus, the current study examined whether contextual information (i.e., the background color), a component of variability which exerts an automatic influence during spaced study, is stored in the memory representation. In doing so, we brought reminding of background color under conscious control during the acquisition phase in Experiment 1, whereas in Experiment 2 reminding of background color was spontaneous during the acquisition phase but brought under conscious control during the retrieval phase. If elements of contextual variability are automatically stored in the memory representation, then reminding accuracy of color background should be similar across experiments and should exert similar influence on final test performance across experiments. In contrast, if background color is only incorporated in the underlying memory trace when attention is directed toward the dimension of the stimulus, then accuracy for color background should be greater in Experiment 1 than Experiment 2 as measured during the encoding and retrieval phases, respectively. In turn, there should be different effects across encoding variability conditions in final test performance in Experiment 1 when compared against Experiment 2.

In pursuing these aims, it is necessary to first introduce the independent and dependent variables in this 2 (Encoding Variability: constant, variable) x 3 (Lag: 0, 3, 10 intervening items) within-participants design. The two independent variables used in the current study were encoding variability and lag. Encoding variability was manipulated via the background color (i.e., red and green) each trial was presented on such that constant encoding occurred when an item was presented twice on the same background color and variable encoding occurred when an item was presented twice on different background colors. Lag was manipulated such that an item's repetition occurred following no intervening items (i.e., Lag 0, massed), three intervening items (Lag 3), or 10 intervening items (Lag 10).

There were four dependent variables assessed during the acquisition phase and two dependent variables assessed during the final test phase. With regard to the acquisition phase, *reminding accuracy* and *reminding latency* were measured with regard to participants' responses to the reminding judgment (i.e., "Was this item presented previously during the study phase?") which was prompted on every trial. Specifically, accuracy was measured as the proportion of correct detections of repetition on an item's second presentation, and reminding latency was the time taken (in milliseconds; ms) to respond to this prompt. *Reminding of encoding variability accuracy* and *reminding of encoding variability latency* were measured as the proportion correct responses and the time it took (in milliseconds) for participants to make the reminding of encoding variability judgment (i.e., "Was this item previously studied on the same or different background color?") which was prompted for every item that participants judged as a repetition. Final test performance was assessed in two ways. First, *cued recall accuracy* was measured as the proportion of correct cued recall trials. *Recognition accuracy* was measured using an intact/rearranged recognition test and was calculated as the proportion of hits (i.e., calling an intact item intact) minus the proportion of false alarms (i.e., calling a rearranged item intact).

Experiment 1

Methods

Hypotheses

Considering *reminding accuracy*, I predict a main effect of lag such that reminding accuracy will decrease as lag between repetitions increases, and a main effect of encoding variability such that reminding accuracy is lower overall for the variable encoding condition compared to the constant encoding condition. I also predict an interaction such that the decrease in reminding accuracy across lag conditions will be larger in the variable encoding condition compared to the constant encoding condition. Moreover, for *reminding response latency*, I predict a main effect of lag such that reminding latency will be longer (i.e., more effortful) for Lag 10 items relative to Lag 3 and massed items. I also predict an interaction such that the increase in response latency across lag conditions will be larger in the variable encoding condition compared to the constant encoding condition.

Considering *reminding of encoding variability accuracy*, I predict a main effect of lag such that accuracy of reminding of encoding variability will decrease with increases in lag. Furthermore, for *reminding of encoding variability response latency,* I predict a main effect of lag such that the time taken to be reminded of encoding variability will increase with increases in lag.

For final test performance, I predict main effects of both lag and encoding variability on both the *cued recall test* and the *recognition test*. The patterns should be the same but may be diminished in size on the recognition test provided the additional retrieval support provided on the recognition test (i.e., retrieval cues for both members of the word pair) compared to the cued recall test (i.e., retrieval cues for the left member of the word pair only). In terms of lag, I expect accuracy to be greater for Lag 10 items relative to Lag 3 and Lag 0 items. With regard to encoding variability, I predict that accuracy will be enhanced for pairs presented in the variable encoding condition relative to pairs presented in the constant encoding condition independently of reminding latency. Finally, I predict an interaction such that encoding variability will benefit final test performance for massed items but not for spaced items.

Participants

Forty-eight participants were recruited for Experiment 1 from introductory psychology courses at Rhodes College (mean age = 18.06). Participants were given research credits toward partial fulfillment of the introductory psychology course requirement.

Design

Experiment 1 utilized a 2 (Encoding Variability: variable, constant) x 3 (Lag: 0, 3, 10), within-participants design. Encoding variability was manipulated by the color of the background (red or green) on which word pairs were presented: In the constant encoding condition, an item was presented twice on the same background color, whereas in the variable encoding condition, word pairs were presented once on a red background and once on a green background. Items were repeated after one of three lags (0, 3, or 10 intervening items).

Stimuli

The stimuli set for this experiment consisted of 83 word pairs. Of these 83 pairs, 60 were critical pairs (i.e., presented twice during the Acquisition Phase) and the remaining 23 were filler pairs (i.e., presented once during the Acquisition Phase). The 60 critical pairs were divided into six sets of ten pairs which were equated on mean word length and HAL frequency (Balota et al., 2007) separately for both cues and targets (for cues and targets, all *p*s > .30). Furthermore, using normed forward associative strength (see Nelson, McEvoy, & Schreiber, 2004) and latent semantic analysis (Landauer, Foltz, & Laham, 1998), mean semantic relationship between cue and target was statistically equivalent across stimulus sets (*M* = .280, *SE* = .117, *p*s > .30). All sets were counterbalanced across lag and encoding conditions such that each set appeared equally often in each combination of lag and encoding variability conditions across participants. Finally, filler items were used to ensure that serial position was equated for first and second presentations across all lag and encoding variability conditions (*p*s > .45)

Procedure. The current study included three distinct phases: acquisition, cued recall, and recognition. During the acquisition phase, participants studied word pairs one at a time to remember for a later memory test. During this phase, to-be-remembered word pairs were displayed on the screen for six seconds each. *Controlled reminding* was utilized during this phase such that participants were asked to make judgments about previous presentations of each word pair (i.e., have you studied this word pair before?) and the previous encoding condition (i.e., was the first presentation of this pair on the same or a different colored background?) on each trial. Following the completion of this phase, participants completed the cued recall phase during which each cue (i.e., left-hand word of each pair) was presented in a random order, and participants were asked to recall the target (i.e., right-hand word of each pair). The cue word was on a black background and displayed until participants responded with a plausible target or ‘I don’t know.’ Following cued recall, participants completed an intact-rearranged recognition test such that word pairs were either intact (i.e., the cue item and target item were paired such that the word pair was identical to that in the acquisition phase) or rearranged (i.e., the cue item of one pair presented in the acquisition phase was paired with the target item of a different pair presented in the acquisition phase), and participants were asked to make a judgment of ‘intact’ or ‘rearranged’ for each presented pair during the recognition phase. Similar to the cued recall phases, each word pair was displayed on the screen until participants responded.

Following completion of the critical memory task, participants completed an independent measure of episodic memory and a demographics questionnaire. Specifically, participants studied a list of 15 words one at a time before completing a standard demographics questionnaire (e.g., How old are you?; How many years of education have you completed?). Following the completion of the questionnaire, participants were given three minutes to freely recall as many of the 15 previously presented items. In anticipating a cross-experiment comparison following Experiment 2, this independent measure was included to provide a comparison of participant groups across experiments, and in turn, this comparison helps ensure that between-experiment differences are not due to different levels of episodic memory ability across samples.

Results

The purpose of the first experiment was to investigate reminding performance and final test performance under the condition of controlled reminding during the acquisition phase. Thus, I first consider accuracy and response latency for the reminding event (i.e., the second presentation) and for the reminding of encoding variability judgment, and then I consider cued recall test performance.

Reminding accuracy. Mean reminding accuracy (i.e., correctly detecting a repetition on an item’s second presentation) is displayed in Figure 3. There are two observations worth noting. First, reminding accuracy was greatest in the massed condition and comparable across the Lag 3 and Lag 10 conditions. Second, within each lag condition, reminding accuracy was comparable between encoding conditions.

Accuracy of reminding was submitted to a 3 (Lag) x 2 (Encoding Variability) repeated measures ANOVA, and the results supported these observations. Specifically, analyses revealed a significant effect of lag, $F(2, 88) = 7.32$, $p = .001$, $\eta^2_p = .14$, and no significant effect of encoding variability nor a significant Lag x Encoding Variability interaction (ps > .65). Follow-up comparisons with Bonferroni correction indicated that accuracy was significantly greater in the massed condition than the Lag 3 and Lag 10 conditions, (ps < .05), and there was no difference between Lag 3 and Lag 10, (ps > .999).

Standardized reminding response latency[1]. Response latency for successful reminding was standardized individually for each participant. Critically, outliers were removed from analysis if they were less than 200 ms, which resulted in removal of .12% of trials, and if the resulting *z* scores exceeded three standard deviations, which resulted in removal of .4% of trials. Mean standardized response latency is presented in Figure 4. As can be seen in Figure 4, standardized response latency differed by lag and by encoding condition such that, for constant encoded items, response latency increased monotonically with increases in lag and, for variable encoded items, response latency

---

[1] The pattern of mean raw response latency was generally consistent with the pattern of mean standardized response latency presented here.

increased from the massed condition to the Lag 3 condition, which was comparable to the Lag 10 condition.

A 2 (Encoding Variability) x 3 (Lag) repeated measures ANOVA supported these observations, as there were significant differences across lag, $F(2, 88) = 29.06$, $p < .001$, $\eta^2_p = .40$, and encoding variability approached significance, $F(1,44) = 3.91$, $p = .054$, $\eta^2_p = .08$. These results were further qualified by a significant Encoding Variability x Lag interaction, $F(2, 88) = 5.50$, $p = .006$, $\eta^2_p = .11$. Follow up ANOVAs revealed a significant effect of lag in both the constant encoding condition, $F(2,88) = 30.47$, $p < .001$, $\eta^2_p = .41$, and the variable encoding condition, $F(2,88) = 11.83$, p < .001, $\eta^2_p = .21$. Follow up comparisons with Bonferroni correction revealed that, in the constant encoding condition, massed items were reminded faster than both Lag 3 and Lag 10 items (*p*s < .001) while Lag 3 items were reminded faster than Lag 10 items ($p < .01$), whereas in the variable encoding condition, massed items were reminded faster than both Lag 3 and Lag 10 items (*p*s < .05) but there was no difference in reminding response latency between Lag 3 and Lag 10 items ($p > .999$). To better understand the influence of encoding variability within each lag condition, targeted *t* tests were conducted. Results revealed significant effects of encoding variability in Lag 0, $t(44) = 2.88$, $p = .006$, and Lag 3 conditions, $t(44) = 2.21$, $p = .032$, but there was no difference in standardized response latency across constant and variable encoding for repetitions in the Lag 10 condition, $t(44) = 1.51$, $p > .10$.

Reminding of encoding condition accuracy. Mean accuracy for reminding of encoding condition is displayed in Figure 5. It is worth noting that reminding of encoding condition accuracy decreased monotonically with increases in lag such that reminding accuracy was greatest in the massed condition.

A 3 (Lag) x 2 (Encoding Variability) repeated measures ANOVA supported these observations, as they revealed a single significant effect of lag, $F(2, 88) = 172.25$, $p < .001$, $\eta^2_p = .80$. All follow-up comparisons were significant such that accuracy was highest in the Lag 0 condition, moderate in the Lag 3 condition, and lowest in the Lag 10 condition (ps < .005). It is important to note that performance in the Lag 0 and Lag 3 condition was significantly above chance (ps < .05), whereas performance in the Lag 10 condition was no greater than chance ($p > .05$).

Reminding of encoding condition response latency. Mean response latency for reminding of encoding condition is displayed in Figure 6. It is worth noting that reminding of encoding condition response latency increased between Lag 3 and Lag 10 for constant encoded items, whereas response latency increased between Lag 0 and Lag 3 for variable encoded items.

Data were submitted to a 3 (Lag) x 2 (Encoding Variability) repeated measure ANOVA. Results revealed a main effect of lag, $F(2, 86) = 9.31$, $p < .001$, $\eta^2_p = .18$, and a main effect of encoding variability, $F(1, 43) = 11.85$, $p = .001$, $\eta^2_p = .22$. These main effects were further qualified by a significant interaction, $F(2, 86) = 4.89$, $p = .010$, $\eta^2_p = .10$.

Cued recall accuracy[2]. Mean cued recall accuracy is displayed in Figure 7. There are two points worth noting in this figure. First, accuracy for both encoding conditions revealed a nonmonotonic, inverted U-shape function that is characteristic of the spacing effect in the cognitive literature. Second, cued recall accuracy was generally higher for items in the variable encoding condition relative to constant encoded items, and this increase in performance is most obvious in the Lag 10 condition, modest in the massed condition, and negligible in the Lag 3 condition.

A 3 (Lag) x 2 (Encoding Variability) repeated measures ANOVA supported these observations, as analyses revealed significant effects of lag, $F(2, 88) = 13.75$, $p < .001$, $\eta^2_p = .24$, and encoding variability, $F(1, 44) = 10.83$, $p = .002$, $\eta^2_p = .20$. Furthermore, the Lag x Encoding Variability interaction approached significance, $F(2, 88) = 2.52$, $p = .087$, $\eta^2_p = .05$. Given that I aimed to assess the benefit of encoding variability *a priori*, follow up ANOVAs were conducted on this interaction. These follow up tests revealed a significant main effect of lag in the constant encoding condition, $F(2,88)$, $p < .001$, $\eta^2_p = .302$, such that performance for Lag 3 items was greater than both massed and Lag 10 items (*p*s < .01) and that performance for massed items was greater than that of performance for Lag 10 items ($p < .03$). No significant effect of lag in the variable encoding condition was observed ($p > .250$).

---

[2] Cued recall was also conditionalized for items successfully reminded as a repetition, as well as for items successfully reminded as repetitions and successfully reminded as constant or variable encoding. Results of the conditional analyses revealed the same pattern of results seen in the unconditionalized data. This is not surprising given the high reminding success observed across all lags and encoding conditions; indeed, 91% of items were included in the conditional analysis under these constraints. The Lag x Encoding Variability interaction also reached significance, $F(2,88) = 3.11$, $p = .049$, $\eta^2_p = .07$. This interaction was further investigated with follow up ANOVAs that revealed a significant effect of Lag in the constant encoding condition, $F(2,88) = 17.87$, $p < .001$, $\eta^2_p = .29$, but not in the variable encoding condition ($p > .450$). Within the constant encoding condition, cued recall performance for lag 3 items exceeded that of both massed items and lag 10 items (*p*s < .01). Furthermore, cued recall accuracy for massed items was greater than cued recall accuracy for lag 10 items ($p < .03$)

Discussion

Findings from Experiment 1 yielded two conclusions with regard to better understanding the mechanism underlying the spacing effect. First, results suggest a limited role for desirable difficulty as a contributor to the spacing effect (see also Maddox et al., in prep). Notably, if one assumes that response latency serves as a proxy for retrieval difficulty, then the pattern observed in reminding response latencies should be replicated in final test performance. This was not the case. With regard to the constant encoding condition, reminding response latencies increased monotonically with increasing lag, but cued recall produced the typical inverted-U shape function. Although the difference between Lags 0 and 3 can be accounted for by increased reminding difficulty, the discrepancy across response latency and final test performance when comparing Lags 3 and 10 cannot be accommodated by the desirable difficulty account. Next consider the variable encoding condition. Response latency increased from massed to spaced conditions, but there was no difference in reminding difficulty across Lags 3 and 10. Nevertheless, there was no difference in final test performance across massed and spaced conditions. Importantly, reminding accuracy was largely successful across lag and encoding conditions, and thus reminding failure cannot mediate the observed dissociation between reminding response latency and final test performance. Simply put, reminding difficulty as operationalized in the current study failed to fully predict long-term memory performance for repeated study in both encoding conditions, and thus the results of the current study cannot be sufficiently accommodated by Benjamin and Tullis' (2010) desirable difficulty and reminding dual-mechanism.

Second, consider the extent to which the current results can be accommodated by the encoding variability account. Based on the results of Verkoeijen et al. (2005), one would predict a benefit of variable encoding for the massed condition, whereas for the spaced conditions one would predict equivalent performance for constant and variable encoded items. Cued recall performance supported these predictions in the massed condition, which revealed a benefit of variable over constant encoding, and in the Lag 3 condition, which revealed no difference between constant and variable encoded items. However, cued recall performance deviates from these predictions in the Lag 10 condition as a benefit of variable encoding over constant encoding was observed. Thus, while encoding variability can accommodate cued recall performance in the massed and Lag 3 condition, the mechanism cannot accommodate the cued recall results for Lag 10.

It is possible that the observed deviation from the predictions of Verkoeijen et al. (2005) in Lag 10 could indicate that the efficacy of the encoding variability mechanism in the current paradigm may be qualified by reminding quality in terms of participant utilization of recollection- and familiarity-based reminding. A reminding event that is based on recollection should include reminding of extraneous details (i.e., background color) when the item is successfully detected as a repetition, whereas a reminding event that is based on familiarity should not include the reminding of these details even if the item is successfully detected as a repetition. Importantly, the use of recollection- and familiarity-based reminding is not mutually exclusive, but rather both bases for reminding can be utilized for an item. Furthermore, given that the current paradigm was not designed to assess the contributions of recollection versus familiarity, discussion of the current results in these terms is largely speculative.

With this in mind, reconsider the benefit of variable over constant encoding in cued recall performance for the Lag 10 condition. For reminding response latency in the

Lag 10 condition, statistically equivalent speed between constant and variable encoded items was observed. Furthermore, reminding of encoding condition was statistically equivalent to chance. These patterns of reminding response latency and reminding of encoding condition accuracy suggest two things in terms of recollection- and familiarity-based reminding for the Lag 10 condition. First, it suggests that the quality of reminding was comparable between encoding conditions, as reminding response latency did not differ between encoding conditions. Second, it suggests a familiarity-basis for reminding such that, while reminding of repetitions was near perfect, recollection of background color did not occur. Taken together, these results suggest that in the absence of recollection-based reminding in the Lag 10 condition, variable encoding conditions may exhibit an implicit influence on final test performance.

Reconsider the results in the massed and Lag 3 conditions, which supported Verkoeijen et al.'s (2005) predictions in terms of encoding variability. For the massed condition, relatively fast reminding response latency suggests a heavy reliance on familiarity relative to recollection as a basis for reminding. However, the observed increase in response latency for variable relative to constant encoded items in the massed condition suggests that variability induced disfluency of familiarity-based reminding such that the change in background color between the first and second presentation of a massed item prompted participants to consciously recollect the background color of the first presentation. On the other hand, for the Lag 3 condition, relatively longer reminding response latency suggests a heavier reliance on recollection as a basis for reminding. However, unlike the Lag 10 condition, reminding of encoding condition accuracy for the Lag 3 condition was statistically above chance. This suggests that the constraints imposed by encouraging looking back was indeed greater at long lags (i.e., Lag 10) relative to short lags (i.e., Lag 3)

Experiment 2

Experiment 2 examined the influence of lag and encoding variability without requiring participants to provide trial-by-trial repetition detection and judgments of encoding condition, and instead these processes were allowed to occur spontaneously throughout the acquisition phase. Nonetheless, these measures were collected when participants completed cued recall in the form of two retrospective judgments (i.e., recollection of reminding). In doing so, I had two specific aims. First, allowing reminding to be spontaneous in the encoding condition may free up cognitive resources that were previously allocated to the intentional processing of contextual components of variability as well as the "looking back" process for those components. Second, removing the constraints of controlled reminding of contextual variability components, which are assumed to exert an automatic influence on encoding when not brought under conscious control, may more fully allow the participants to encode and reap the benefits of structural and descriptive variability. If participants can engage in controlled processing of variable components that will be more useful retrieval cues (e.g., inter-stimuli associations) relative to low utility cues (i.e., background color) encoded under more automatic influences, then spontaneous remindings may be more recollection-based. This may be particularly true for the Lag 10 condition because controlled reminding prompts looking back, which is expected to take longer and thus impose more constraints on controlled processing at this lag relative to the massed and Lag 0 conditions. In turn, it is predicted that cued recall performance will increase monotonically or may approach asymptote as lag increases. Moreover, to the extent that participants may engage in processing of structural and descriptive components of variability without allocating controlled attention to the processing of the contextual component of background color –

which is otherwise an unhelpful cue for reminding and for long-term cued recall performance – there should be no effect of encoding variability.

Methods

Participants

36 participants were recruited for Experiment 1 from introductory psychology courses at Rhodes College (mean age = 18.98). Participants were given either credit for partial fulfillment of the introductory psychology course requirement or $10 for compensation.

Materials, Design, and Procedure

The methods for Experiment 2 were identical to Experiment 1 with two exceptions. First, Experiment 2 utilized *spontaneous reminding* during the acquisition phase. In other words, participants were not prompted a judgment concerning previous repetitions, nor previous encoding conditions, of the presented word pairs. Instead, participants were asked to make a filler judgment, specifically concerning the animacy of each word pair (i.e., Are either words in the pair animate?). Participants were instructed to respond “yes” if either the cue or the target were animate, and “no” if both words in the pair were inanimate.

Second, Experiment 2 utilized *controlled recollection of reminding* during the retrieval phase. Specifically, during the cued recall portion of the retrieval phase, participants were asked to make two judgments that encouraged reminding of each word pair as it was presented during the acquisition phase. The first judgment pertained to repetitions of items (i.e., was this word pair presented once or multiple times during the acquisition phase?), and the second judgment pertained to the encoding condition of items (i.e., was this word pair presented twice on the same colored background or different colored backgrounds during the acquisition phase?). Note that this second judgment only occurred if participants responded that the word pair in question had indeed been presented twice during the acquisition phase. If they responded that it had only been presented once, they proceeded to the next pair.

Results

Cued recall accuracy. Mean accuracy for cued recall is displayed in Figure 8. It is worth noting that cued recall performance was lowest for massed items and comparable between Lag 3 and Lag 10 items.

A 3 (Lag) x 2 (Encoding Variability) repeated measures ANOVA revealed a main effect of lag, $F(2,78) = 7.17$, $p = .001$, $\eta^2_p = .15$. This effect was further investigated utilizing follow up comparisons with Bonferroni correction, and analyses revealed that mean cued recall performance for massed items was significantly lower than that of Lag 3 and Lag 10 items ($p < .02$), whereas Lag 3 and Lag 10 items did not differ ($p > .999$). No significant effect of encoding variability was found ($p > .750$), nor was the Lag x Encoding Variability interaction significant ($p > .150$).

Recollection of reminding. Mean recollection of reminding accuracy for repeated items is displayed in Figure 9. Most notably, recollection of reminding was least successful (i.e., a higher proportion of responses indicating that a repeated item had not been repeated) in the massed condition, whereas Lag 3 and Lag 10 recollection of reminding performance was comparable.

The results of a 3 (Lag) x 2 (Encoding Variability) repeated measures ANOVA supported this observation, as it revealed a main effect of lag, $F(2,66) = 15.40$, $p < .001$, $\eta^2_p = .32$, that when further analyzed using Bonferroni comparisons revealed differences in performance between the massed and spaced conditions such that recollection of reminding was lower for massed items than it was for Lag 3 and Lag 10 items (*p*s <.01), whereas recollection of reminding performance did not differ between the Lag 3 and Lag 10 conditions ($p > .40$). Furthermore, 3 (Lag) x 2 (Encoding Variability) repeated measures ANOVA revealed a potential effect of encoding variability that approached significance, $F(1,33) = 3.75$, $p = .062$, $\eta^2_p = .10$, but the Lag x Encoding Variability interaction did not reach significance ($p > .10$).

Recollection of encoding variability. Figure 10 represents mean accuracy of recollection of encoding variability. As depicted in the figure, mean recollection of encoding variability accuracy decreases monotonically for constant encoded items and increases monotonically for variable encoded items.

The results of a 3 (Lag) x 2 (Encoding Variability) repeated measures ANOVA supported these observations, as results revealed no significant main effect of lag ($p > .950$) nor encoding variability ($p > .250$), but the Lag x Encoding Variability interaction reached significance, $F(2,54) = 3.22$, $p = .048$, $\eta^2_p = .10$. Follow up ANOVAs suggested no main effect of lag for constant encoded items ($p > .250$), but a main effect of lag that approached significance for variable encoded items, $F(2,62) = 2.71$, $p = .074$, $\eta^2_p = .08$. Follow up comparisons with Bonferroni correction, however, revealed no significant differences between lag conditions for variably encoded items (*p*s > .250).

Discussion

The results of Experiment 2 supported my hypotheses. Specifically, removing the reminding judgments from the encoding phase facilitated the binding of repeated items to contextual components during the encoding phase as is evidenced by an overall increase in cued recall performance, though one must be cautious to draw strong conclusions on this front given that this is a between-groups comparison. Moreover, removing these judgments during the encoding phase and allowing controlled attention to be allocated to encoding of structural and descriptive contextual components may have been more useful in producing robust traces for later retrieval. Specifically, the asymptotic increase observed in cued recall performance for Experiment 2 when participants were not required to utilize controlled attention to encode the background color during acquisition suggests that performance at long lags may have been hindered by “looking back” task demands in Experiment 1 that emphasized memory and reminding for background color. Furthermore, given that this asymptotic increase was observed when recollection of encoding variability was equal to chance in all lag conditions, the results of Experiment 2 also support my prediction that bringing reminding of contextual components under conscious control only appears to facilitate encoding when recollection-based reminding is available. However, structural and descriptive components likely will be able to facilitate encoding even in the absence of recollection-based reminding, and indeed, this appears to be the case in Experiment 2.

One variable thus far unaccounted for in this interpretation of the results of Experiment 2 is the use of animacy judgments during the acquisition phase. While such judgments do not require looking back, as intended, they may facilitate deeper processing of the semantic relationship between the cue and target within a pair, and as such may introduce a confounding element of encoding variability (i.e., descriptive components of variability). However, this potential confound is not expected to produce systematic effects in the current results for two reasons. First, controlled encoding of descriptive components of variability is not expected to diminish the influence of the contextual component of variability, background color, that was manipulated across trials. This is because contextual components of variability exert an automatic influence during study, and thus can co-occur with other, more controlled encoding of variability. Second, word pairs were counterbalanced across lag and encoding variability conditions, and as such animacy judgments were made for all pairs in each lag and encoding variability condition. Thus, while there is not sufficient evidence to deny an effect of animacy judgments on the current results, there is no reason to believe that the animacy judgments made a significant contribution to the increase in performance observed in the Lag 10 condition.

Cross Experiment Comparisons of Cued Recall Performance

To further assess the long-term memory effects of automatic and controlled remindings as well as the open versus constrained processing of encoding variability during acquisition, cued recall performance was compared across Experiments 1 and 2.

Figure 11 depicts mean cued recall accuracy as a function of lag, encoding variability, and experiment. There are two observations to note. First, performance is higher in Experiment 2 compared to Experiment 1. Second, the benefit of lag and encoding variability clearly differs across experiments. Cued recall performance increases and approaches asymptote across lag conditions for constant and variable encoding conditions in Experiment 2. In contrast, Experiment 1 performance is comparable across lag conditions in the variable encoding condition, and the typical inverted-U shape function relating lag to final test performance is observed in the constant encoding condition.

These observations were supported by the results of a 2 (Experiment) x 2 (Encoding Variability) x 3 (Lag) mixed-factors ANOVA. All main effects were significant, *p*s < .01, and were further qualified by several interactions. First, the Experiment x Lag interaction was significant, $F(2, 166) = 8.61$, $p < .001$, $\eta^2_p = .09$. As has been previously discussed for each of the individual experiments, the lag effect in Experiment 1 was significant ($p < .001$) and reflected the inverted U-shape function in which all conditions were significantly different from one another (*p*s < .05). For Experiment 2, the lag effect was also significant ($p = .001$) and reflected significantly improved performance for Lag 3 and Lag 10 conditions relative to the Lag 0 condition (*p*s < .05). Second, the Experiment x Encoding Variability interaction was significant, $F(1, 83) = 5.48$, $p = .022$, $\eta^2_p = .062$. The encoding variability effect was significant in Experiment 1 ($p = .022$) but was not significant in Experiment 2 ($p > .75$).

General Discussion

The results of the current study provide insight into the joint influence of desirable difficulty, encoding variability, and reminding. The implications of the present results will be considered for each mechanism in turn. First, desirable difficulty plays a limited role in producing a benefit of spaced study, and this is evidenced by the results of reminding latency and cued recall performance in Experiment 1. Specifically, reminding latency revealed a monotonic increase in response latency with increases in lag for constant encoded items, and an asymptotic increase in response latency with increases in lag for variable encoded items. Assuming response latency serves as a proxy for reminding difficulty, one would expect a monotonic increase in cued recall performance for constant encoded items and an asymptotic increase in cued recall performance for variable encoded items. This was not observed. As such, the reminding difficulty mechanism cannot adequately account for all of the observed results in Experiment 1.

The second implication of the current research is that encoding variability also has trouble accommodating the benefit of spaced study observed in long-term memory performance. Specifically, the extent to which encoding variability can facilitate a long-term memory benefit of spaced study may be modulated by the quality (i.e., the extent to which participants can utilize recollection-based reminding) of the reminding event. Joint consideration of acquisition phase performance (i.e., reminding accuracy and response latency, as well as reminding of encoding variability accuracy and response latency) suggested that encoding variability in the massed condition may disrupt reminding fluency such that it triggers an event of change detection, which in turn should lead to greater processing of variable over constant encoded items under massed conditions.

Analyses of the Lag 3 acquisition performance suggested a role of recollection-based reminding for constant encoding items and more familiarity-driven reminding for variable encoding items. The difference in bases for reminding may have contributed to equivalent final test performance such that recollection-based reminding can compensate for the lack of encoding variability for constant items in the current experiment. Finally, at Lag 10, reminding of encoding condition was not significantly different from chance, which suggests reminding based on familiarity at this lag. In turn, without clear recollection of the original learning event, any variability that may operate automatically during acquisition (i.e., contextual components like background color) may confer some minimal benefit. Thus, the current results suggest that a benefit of variable encoding over constant encoding seems contingent upon the quality of the reminding event.

In light of this interpretation of the results of Experiment 1, it could also be the case that bringing the spontaneous encoding of contextual components under task control had negative influences on reminding quality. Specifically, the controlled reminding of the contextual component of background color requires “looking back” from the second presentation of an item to the first presentation of that item. Controlled reminding in such a fashion can disadvantage the encoding of to-be-learned material at long lags in two ways. First, the cognitive resources required for looking back become more taxing as the interval between the first and second presentation of an item increases. Second, fewer cognitive resources are then available to process more efficient components of the reminding event, such as structural or descriptive components. If such detriments of bringing the encoding of contextual components under conscious control can indeed explain the cued recall performance results of Experiment 1, one would expect the decrease in performance from Lag 3 to Lag 10 to be diminished when the encoding of contextual components is left to spontaneous reminding. This is exactly what was observed in Experiment 2. This conclusion was further supported by the cross-experiment interaction, which revealed a significant effect of encoding variability under controlled reminding conditions in Experiment 1 but no effect of encoding variability under the spontaneous reminding conditions of Experiment 2.

The third implication of the current research concerns of the way in which reminding contributes to the benefit of spaced study. Critically, controlled reminding is assumed to be more successful at longer lags than is reminding that occurs spontaneously. The current study suggests that controlled reminding can overreach and actually harm long-term performance. Specifically, the utilization of controlled reminding must be prioritized in terms of high utility cues. In other words, bringing reminding under conscious control can facilitate a long-term benefit of spaced study assuming that it is used for the reminding of relevant cues. Conversely, bringing reminding under conscious control can facilitate detriments of spaced study to the extent that it ignores relevant cues in favor of low utility cues. This conclusion is evidenced both by the decrease in cued recall performance observed between Lag 3 and Lag 10 of Experiment 1, as well as the benefit of spontaneous reminding in Experiment 2 over the controlled reminding of contextual components in Experiment 1. Indeed, removing the controlled reminding judgment regarding encoding variability condition while retaining the repetition detection judgment is sufficient for producing the pattern of results observed in Experiment 2 (Kauffman & Maddox, 2016).

Taken together, the results of two experiments and a cross-experiment analysis suggest that neither Benjamin and Tullis' (2010) dual-mechanism combining desirable difficulty and reminding nor Greene's (1989) dual-mechanism combining encoding variability and reminding are sufficient in fully accommodating the benefits to long-term memory performance as a result of spaced study. The current results suggest a more complex relationship between reminding and encoding variability than previously posited. Specifically, reminding does not only need to be successful to produce a spacing effect under variable encoding conditions – the nature of reminding (i.e., recollection-based versus familiarity-driven) must also be considered. As such, future research should focus on identifying factors of reminding that may facilitate long-term memory performance under variable encoding conditions, and such research could begin by manipulating controlled versus spontaneous reminding within participants. Indeed, the current results can inform the benefit of spaced study in specific domains such as the classroom, job site training, and cognitive therapies for healthy older adults. Furthermore, given that the current results replicated the findings of Verkoeijen et al. (2005) that introducing contextual encoding variability is most beneficial under massed conditions, the current results also have the potential to facilitate learning in domains which benefit most from massed practice, such as desensitization treatment for phobias (Dua, 1972).

TABLES, ILLUSTRATIONS, AND FIGURES

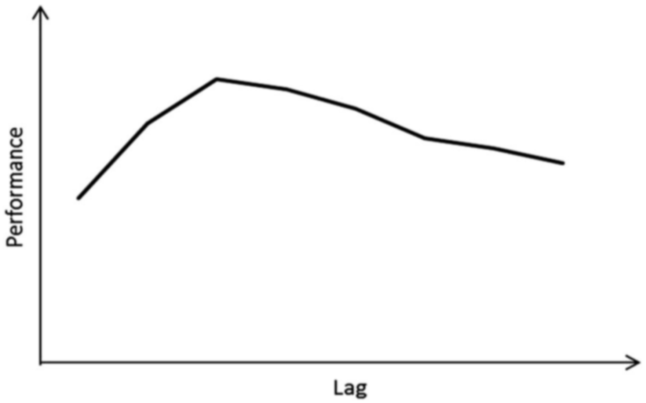

Figure 1. Inverted U-shape function relating lag and long-term memory performance. As lag increases from zero (or massing), memory performance increases until an optimal lag is reached. At this point, memory performance peaks before decreasing with increases in lag.

Item	**Presentation**	**Lag**
plane	1	Lag 10
piano	1	
cards	1	
dice	1	Lag 3
fruit	1	Lag 0
fruit	2	Lag 0
table	1	
dice	2	Lag 3
dog	1	
spider	1	
keys	1	
plane	2	Lag 10

Figure 2. Filler items will only be presented once, but critical items will be presented twice. Critical items will be presented with a lag of either 0 items, 3 items, or 10 items. For example, "fruit" is presented twice with 0 items between each presentation (Lag 0, or massed). "Dice" is presented twice with 3 items between each presentation (Lag 3), and plane" is presented twice with 10 items between each presentation (Lag 10).

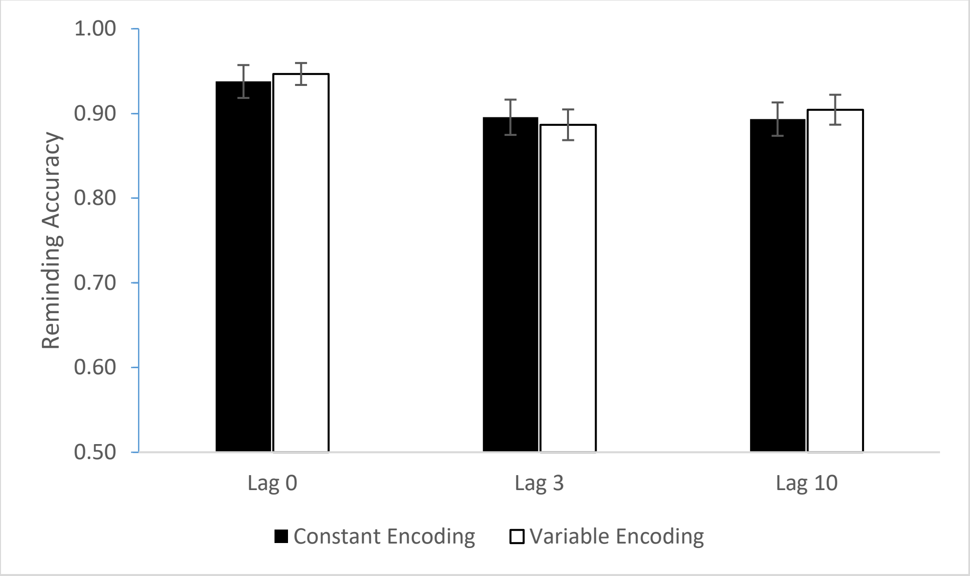

Figure 3. Mean reminding accuracy for detection of repetitions during Experiment 1 acquisition phase as a function of lag and encoding condition.

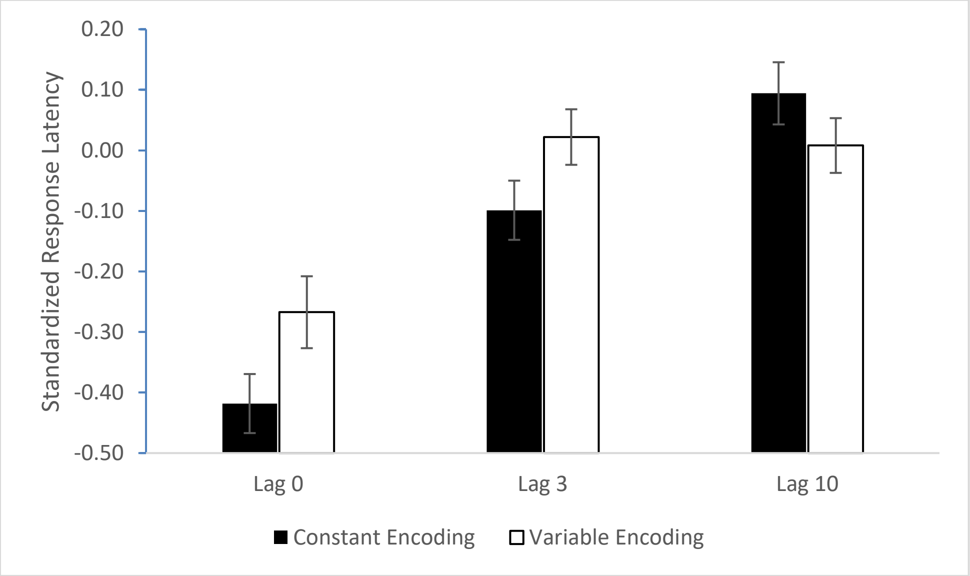

Figure 4. Standardized reminding response latency during Experiment 1 acquisition phase as a function of lag and encoding variability.

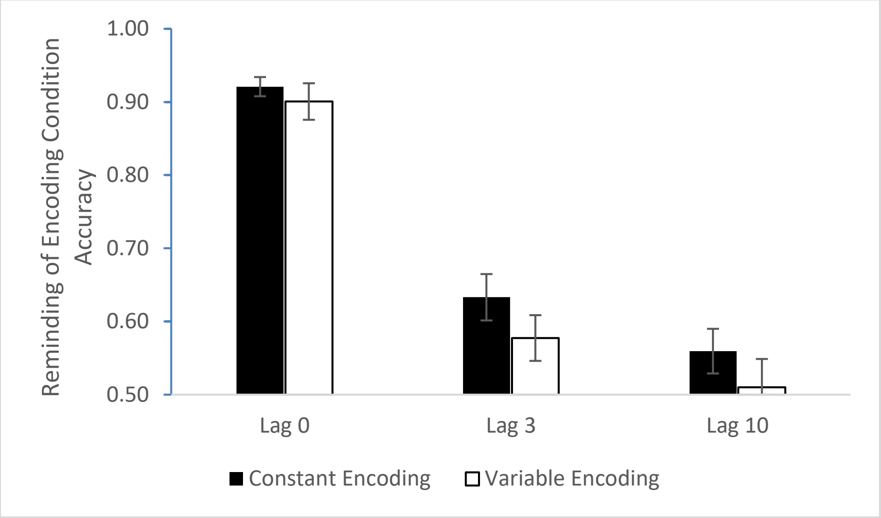

Figure 5. Mean reminding accuracy for detection of encoding variability during Experiment 1 acquisition phase as a function of lag and encoding variability.

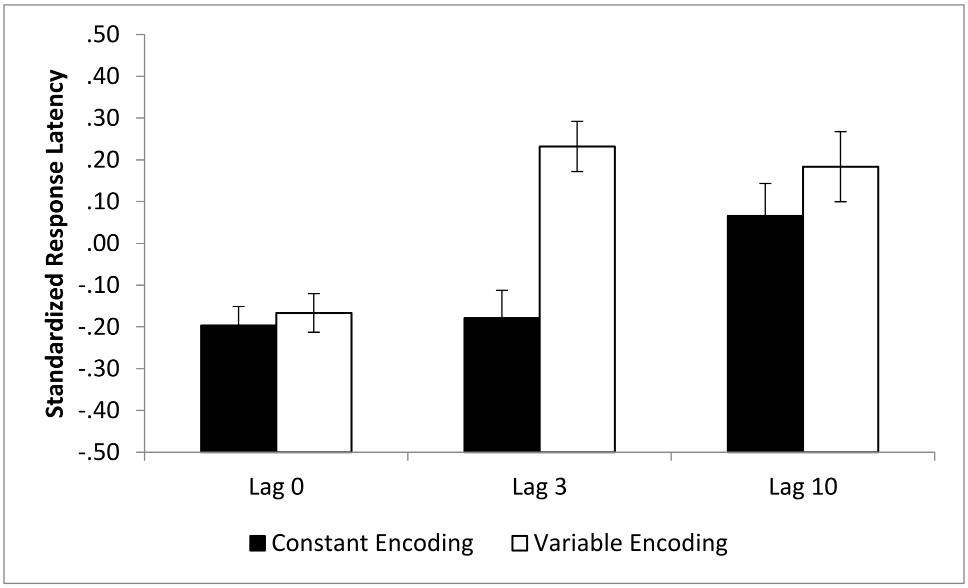

Figure 6. Mean standardized response latency for detection of encoding variability during Experiment 1 acquisition phase as a function of lag and encoding variability

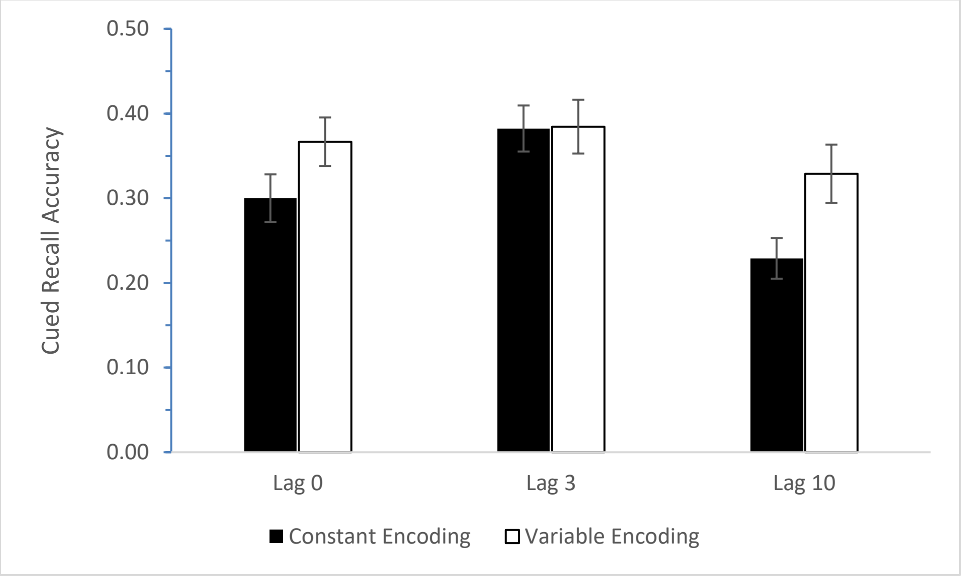

Figure 7. Mean cued recall accuracy during Experiment 1 retrieval phase as a function of lag and encoding variability.

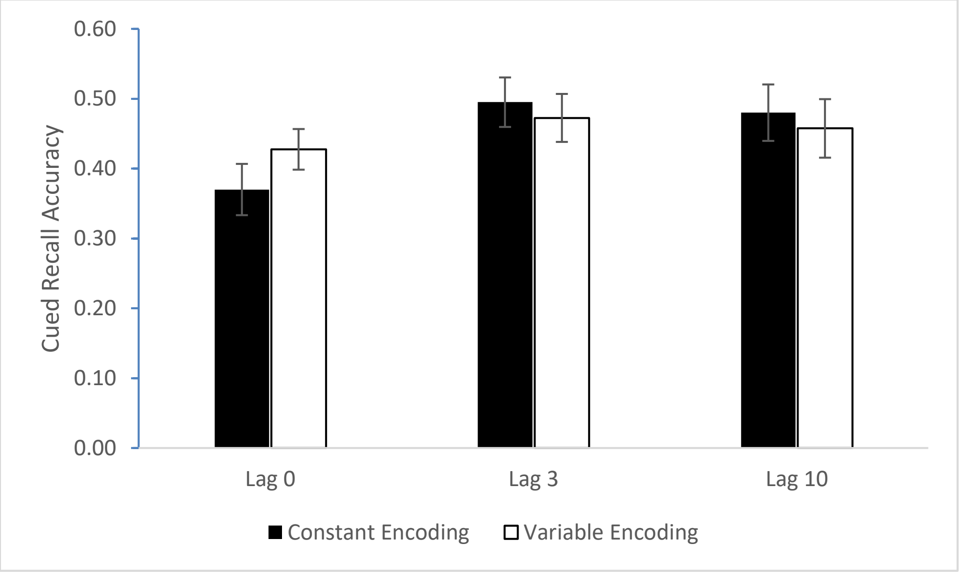

Figure 8. Mean accuracy for cued recall during Experiment 2 retrieval phase as a function of lag and encoding condition.

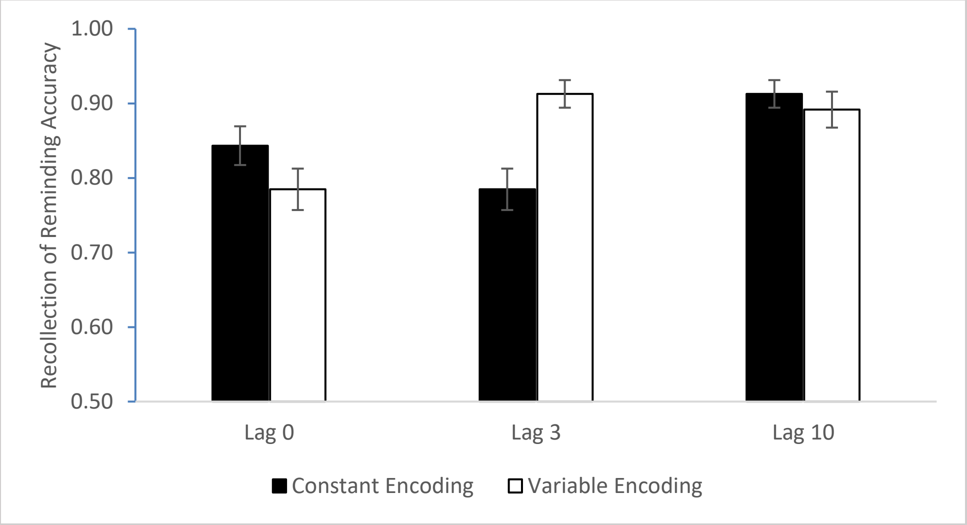

Figure 9. Mean accuracy for recollection of reminding during Experiment 2 retrieval phase as a function of lag and encoding condition.

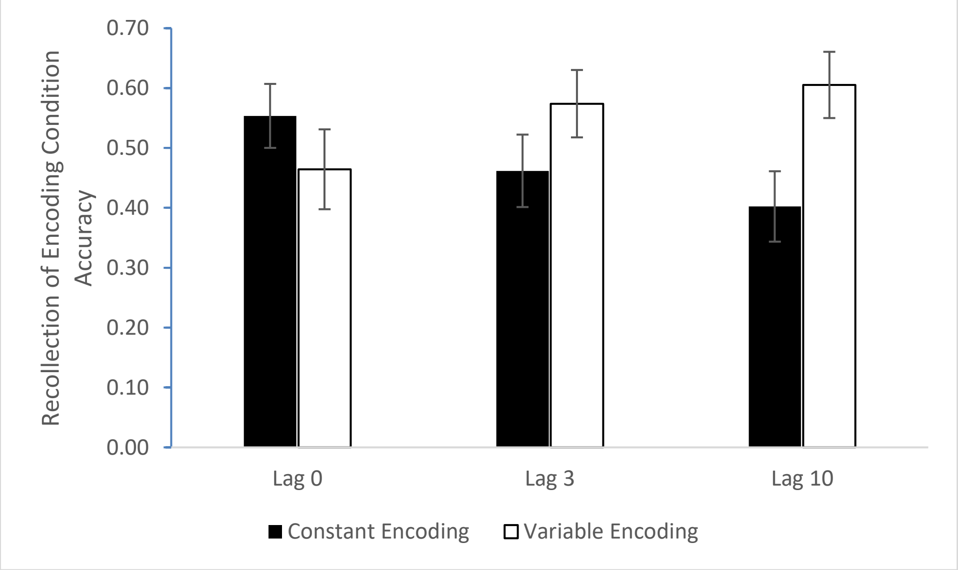

Figure 10. Mean accuracy for recollection of encoding variability during Experiment 2 retrieval phase as a function of lag and encoding condition.

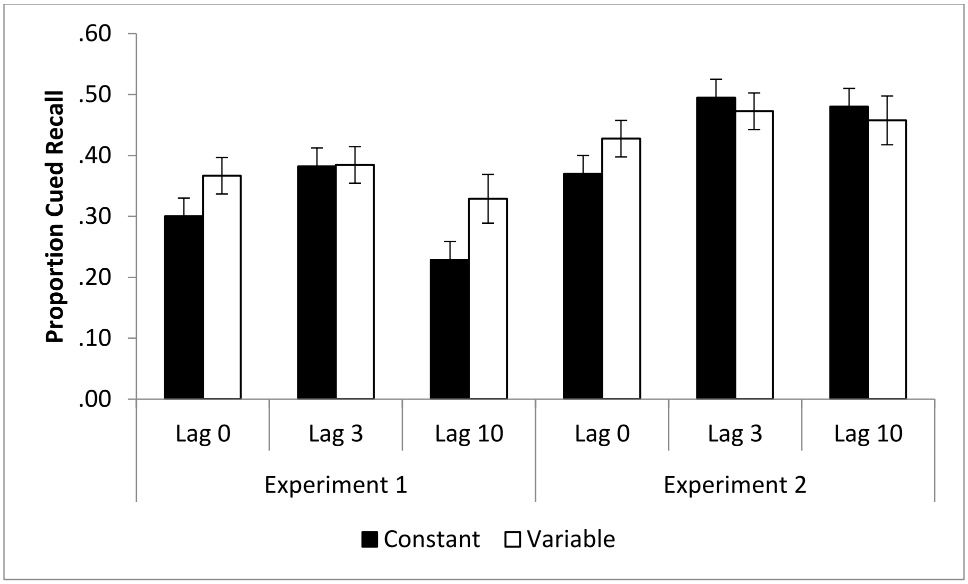

Figure 11. Mean accuracy for cued recall as a function of lag, encoding variability, and experiment.

REFERENCES

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.

Balota, D.A., Yap, M.J., Cortese, M.J., Hutchison, K.A., Kessler, B., Loftis, B., Neely, J.H., Nelson, D.L., Simpson, G.B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, 39, 445-459.

Benjamin, A. S. & Tullis, J. G. (2010). What makes distributed practice effective? *Cognitive Psychology, 61*, 228-247

Bjork, R.A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A.P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185-205). Cambridge, MA: MIT Press.

Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology*. 10 (1): 12–21

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.

Ebbinghaus, H. (1885). *Uber das Gedachtnis*. New York, NY: Dover.

Estes, W. K. (1955). Statistical theory of spontaneous recovery and regression. *Psychological Review*, 62, 145–154.

Gartman, L. M., & Johnson, N. F. (1972). Massed versus distributed repetition of homographs: A test of the differ-ential-encoding hypothesis. *Journal of Verbal Learning and Verbal Behavior*, 11, 801–808.

Greene, R,L. (1989). Spacing effects in memory: Evidence for a two-process account. *Journal of Experimental Psychology: Learning, Memory and Cognition, 15*,371-377.

Glenberg, A. M. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Memory & Cognition*, 7, 95–112. doi:10.3758/bf03197590.

Hintzman, D. L. (1974). Theoretical implications of the spacing effect. In R. L. Solso (Ed.)*, Theories in cognitive psychology: The Loyola symposium* (pp. 77–97). Potomac, MD: Erlbaum.

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

Hintzman, D.L. (2010). How does repetition affect memory? Evidence from judgments of recency. *Memory & Cognition*, *38,* 102-115.

Jacoby, L. L. (1974). The role of mental contiguity in memory: Registration and retrieval effects. *Journal of Verbal Learning and Verbal Behavior*, 13, 483–496.

Kauffman, Z.K., & Maddox, G.B. (2016). In accordance with Margaret Ruffin Hyde Award.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). Introduction to Latent Semantic Analysis. *Discourse Processes*, 25, 259-284.

Madigan, S. A. (1969). Intraserial repetition and coding processes in free recall. *Journal of Verbal Learning and Verbal Behavior*, 8, 828–835.

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*. doi: 10.1080/20445911.2016.1181637

Maddox, G.B., Pyc, M.A., Gatewood, J., Kauffman, Z., & Schonhoff, A. (2015). Examining the contribution of desirable difficult remindings to the spacing effect. Submitted for review.

Melton, A.W. (1967). Repetition and retrieval from memory. *Science*, *158*, 532.

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

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

Peterson, L. R.; Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology*. 58 (3): 193-198

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 Behavior*, 17, 669–680.

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

Underwood, B.J. and Postman, L. (1960). Extra-experimental sources of interference in forgetting. *Psychological Review*, 67, 73-95

Verkoeijen, P. P., Rikers, R. M., & Schmidt, H. G. (2004). Detrimental Influence of Contextual Change on Spacing Effects in Free Recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(4), 796-800.

Wahlheim, C. N., Maddox, G. B., & Jacoby, L. L. (2014). The role of reminding in the effects of spaced repetitions on cued recall: Sufficient but not necessary. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 40, 94–105. doi:10.1037/a0034055.

Wixted, J. T., & Carpenter, S. K. (2007). The Wickelgren Power Law and the Ebbinghaus Savings Function. *Psychological Science*, 18(2), 133-134.