Is variability appropriate? Encoding Variability and Transfer-Appropriate Processing

Jefferson De Jesus Salan

Thesis submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Master of Science
In
Psychology

Rachel A. Diana, Chair
Vanessa Diaz
Anthony Cate

May 22, 2020
Blacksburg, VA

Keywords: encoding variability, context variability, transfer-appropriate processing, episodic memory

Is variability appropriate? Encoding Variability and Transfer-Appropriate Processing

Jefferson De Jesus Salan

ABSTRACT

Transfer-appropriate processing (TAP) proposes that retrieval success is based on the match between processing at encoding and retrieval. We propose that the processing described by TAP determines the contextual cues that are encoded with an event. At retrieval, the presence or absence of contextual cues matching the encoding cues will influence success. To implement these principles as a strategy to improve memory, the nature of future retrieval processing or cues must be known during encoding. As this is unlikely in real-world memory function, we propose that increased encoding variability – increasing the range of encoded cues – increases the *likelihood* of TAP when the retrieval scenario is unknown. The larger the set of encoded cues, the more likely those cues will recur during retrieval and therefore achieve TAP. Preliminary research in our lab (Diana, unpublished data) has found that increased encoding variability improves memory for item information in a novel retrieval context. To test whether this benefit to memory is due to the increased likelihood of TAP, the current experiment compared the effects of encoding variability under conditions that emphasize TAP to conditions that reduce TAP. We found main effects of encoding variability and TAP, but no interaction between the two. Planned comparisons between high and low variability encoding contexts within matching and non-matching retrieval contexts did not produce a significant difference between high and low variability when encoding-retrieval processing matched. We conclude that further studies are necessary to determine whether encoding variability has mechanisms that benefit memory beyond TAP.

Is variability appropriate? Encoding Variability and Transfer-Appropriate Processing

Jefferson De Jesus Salan

GENERAL AUDIENCE ABSTRACT

It is well accepted within the episodic memory literature that successful memory retrieval is often driven by context cues. Specifically, the cues that are stored with the memory of the event. To develop a better understanding of how episodic memory works, we must understand how manipulating context cues changes memory performance. One way to investigate the effects of context manipulation is using encoding variability, which refers to the amount of variability (i.e., change) in context cues from one repetition of an item or event, to the next. Preliminary research in our lab (Diana, unpublished data) has found that increased encoding variability improves memory retrieval in a novel context, but it is unclear why this is the case. We proposed that the mental processing described by transfer-appropriate processing (TAP) – a principle stating that memory retrieval success is determined by the match, or overlap, between the mental processing at encoding (i.e., memory formation) and memory retrieval – determines the contextual cues that are stored with the memory at encoding. We hypothesized that encoding variability works even when TAP has already been achieved by matching the processing and cues at encoding to those at retrieval. Alternatively, we hypothesized that encoding variability works by specifically achieving TAP, so that encoding variability is only helpful when the encoding and retrieval contexts do not match. Results indicated partial support for the alternative hypothesis, suggesting that encoding variability works by achieving TAP. However, these results were not sufficiently conclusive, and it is likely that there are other mechanisms that allow for encoding variability to improve memory. This study establishes the groundwork for future work examining encoding variability and its effects on memory.

Table of Contents

Introduction

Episodic memory is the memory of personally experienced events and it includes information indicating the context in which those events occur (Tulving, 1972; Tulving, 1985). Decades of research have been devoted to studying the effects of encoding and retrieval context on episodic memory success. The encoding context is determined by the initial experience of the event as it is being stored into memory (encoding) and can include information about the temporal position of the event, the spatial location of the event, and frame of mind during the event. Similarly, the retrieval context includes the temporal, spatial, and mental states that are occurring when the encoded event is brought back into conscious awareness (intentionally or not). One of the reasons for the continued interest in encoding and retrieval contexts is that retrieval is thought to be driven by contextual cues, and the effectiveness of those contextual cues is thought to rely both on the relationship of the cues to the encoding context (e.g., Tulving & Thomson, 1973) and on their properties at the time of retrieval (e.g., Nairne, 2002).

Experimental manipulations have demonstrated that the relationship between encoding context and retrieval context is a crucial element of memory success. One such manipulation led to the proposal of the transfer-appropriate processing (TAP) framework, wherein the retrieval of word stimuli was shown to improve when the processing required by the encoding task was reinstated during the recognition test (Morris, Bransford, & Franks, 1978). This framework intuitively coincided with the encoding specificity principle, which states that the contextual elements present at encoding, and the operations performed on those elements (i.e., processing), determine the effectiveness of cues in the retrieval environment (Tulving & Thomson, 1973). As a variant on the notion of encoding-retrieval match - the notion that memory benefits from the overlap between the information in the encoded memory representation and the information in the retrieval environment - the TAP framework's novel contribution was not that the encoding and retrieval contexts had potential benefits for retrieval success (see Godden & Baddeley, 1975), but that the benefits could be induced by matching the task-dependent processing demands, an element of the internal cognitive context. The implications of TAP for the effectiveness of encoding strategies will be discussed shortly.

In addition to direct context manipulations, other research shows that repeated exposures to an item, such as a word or picture, will lead to improved memory for the item. However, this effect can vary depending on the amount of time or number of intervening items between each exposure (Madigan, 1969; Melton, 1967). These effects, broadly referred to as distributed practice effects, have long been the subject of memory research because of their robust nature and their clear practical implications for students and learning in general. One of the long-standing, though disputed, explanations for distributed practice effects suggests that the benefits of spacing item repetitions - increasing the time or number of intervening items in between target item exposures - may be due to the degree of variability in encoding for each of the target item repetitions (e.g. Bower, 1972; Lohnas, Polyn, & Kahana, 2011; Melton, 1970). This account of distributed practice effects, aptly named encoding variability, places an emphasis on the differences in temporal context from one item exposure to the next, implying that the benefits may be attributed to the increased range of contextual cues that are stored with the item representation when repetitions are spaced apart (Melton, 1970).

If encoding variability does in fact lead to more cues stored in the memory trace, then its function, both as a mechanism for distributed practice effects and as a potential method for improving retrieval success, might be conceptually understood in the context of encoding-retrieval match and TAP. In the case of distributed practice, the cognitive and temporal context cues added via each item repetition increase the likelihood that the retrieval context will contain enough elements of the encoded memory trace to evoke successful retrieval. In other words, encoding variability potentially increases the likelihood of a match between encoding and retrieval contexts. If such is the case, then one would expect that increasing the variability of encoding contexts across repetitions of an item would yield effects similar to those of distributed practice. The TAP framework would predict similar results if the encoding context variability is primarily driven by differences in processing at each item repetition, suggesting that encoding variability might also improve retrieval success by inducing TAP. Encoding context variability should act analogously to spacing, resulting in mnemonic benefits for items encoded in variable contexts as compared to those encoded in similar or identical contexts across repetitions.

In the following sections I will briefly review the relevant literature pertaining to encoding variability and TAP. I will present evidence for encoding variability - from within the distributed practice literature and outside of it - as a potentially beneficial encoding strategy. I will then outline and describe the current experiment, conducted in an effort to answer two critical questions, with the second contingent on the first: 1. Can encoding variability be induced to improve retrieval success in recognition memory? 2. If so, does encoding variability yield memory benefits because it increases the likelihood of TAP or for some other reason? Hypotheses will be couched in the crucial role of context in episodic memory and the problem that retrieval context is often not known in advance when memory retrieval is desired or necessary.

Transfer-Appropriate Processing

Developed as a response to the levels of processing framework (Craik & Lockhart, 1972), the concept of transfer-appropriate processing introduced the notion that successful retrieval relies on how well the processing of an item (a word) at retrieval *matches* the processing of that item at encoding, rather than simply how *meaningfully* the item is processed at encoding, as claimed by the levels of processing framework (Morris, Bransford, & Franks, 1977). To show this the authors used an incidental memory procedure in which they presented target words with orienting tasks that evoked either semantic or phonetic processing. The semantic task required the participants to determine whether the target word belonged as part of a sentence, whereas the phonetic task required participants to determine whether the target word rhymed with the last word of a sentence. At test, some participants were given a standard recognition test whereas others were given a rhyming recognition test, wherein they were presented with the rhyme of a target word. The critical finding, replicated across three experiments, was that participants performed better on the rhyming recognition test when the words were acquired under the phonetic task as opposed to the semantic task. This finding challenged the levels of processing notion that semantic (or deep) processing was always superior, while simultaneously establishing the importance of the match between processing at encoding and processing at retrieval for retrieval success.

Further research on the TAP framework demonstrates the importance that encoding-retrieval processing match holds in improving retrieval success. One study replicated the original results leading to the TAP framework and challenged the notion that semantic encoding always enhances recollection (Mulligan & Pickelsimer, 2012). In dual process theories of recognition memory (for a review, see Yonelinas, 2002), recollection is the process primarily responsible for the retrieval of context details whereas familiarity is the process responsible for recognition of individual stimuli. Using two methods for measuring recollection, the remember-know (Tulving, 1985) and process-dissociation (Jacoby, 1991) procedures, Mulligan and Pickelsimer (2002) demonstrated that semantic encoding does not always enhance recollection above perceptual forms of encoding, but rather that the benefits of the encoding process to recollection are contingent upon matching processing at retrieval. Their findings coincide with recent theories that suggest that cognitive context (which includes cognitive processing) is a component of episodic memory context (e.g., Diana, Yonelinas, & Ranganath, 2007). If recollection is memory for the context details of an event, and cognitive processing is an element of those details, then recollection must clearly benefit from TAP, yielding better overall memory for the event.

One of the key issues to be confronted when considering the benefits of TAP outside of a laboratory setting is that the processing and overall context that occur at retrieval are often unknowable in advance. As an example, one might consider a college student preparing for an exam. The form of the question, its wording, and the connections to other concepts are all factors that influence the likelihood of retrieving the correct information on an exam, but the student does not typically know those details in advance of the exam. Therefore, studying in a way that matches the test context specifically is not possible. As such, it should be beneficial for the student to prioritize study efforts that improve the probability of reinstating the context at test. Evidence suggests that encoding variability might be one strategy that facilitates this effort.

Encoding Variability

The encoding variability hypothesis was borne out of work in verbal paired-associate (i.e., stimulus-response or A-B) list learning (Martin, 1968). At the time, there were two empirically interesting findings: 1. Stimulus trigrams (i.e., three-letter compounds) high in meaningfulness (e.g., “CAT”) facilitate learning for paired items (e.g., numbers) in individual lists (e.g. Martin, 1968; Task 1) as compared to trigrams low in meaningfulness (e.g. “RYX”), and 2. Stimulus trigrams low in meaningfulness do not result in proactive interference when learning is to be transferred to a second list with new responses (i.e., changing A-B to A-C) (e.g. Martin, 1968; Task 2). The hypothesis attributed these effects to an inverse relationship between stimulus meaningfulness and encoding variability and posited that low-meaningfulness trigrams, lacking a semantic meaning and composed of separate elements, are likely encoded in a different manner at each learning trial. Thus, low-meaningfulness stimuli lead to slower paired associate learning (due to the variability of the encoded representation) but are more easily associated with new responses due to the flexibility that this variability affords. Much of the immediately ensuing empirical work sought to test this hypothesis with respect to meaningfulness effects and paired-associate transfer situations (e.g. Butler & Merikle, 1970; Merryman & Merryman, 1971; Postman & Stark, 1971; Williams & Underwood, 1970), but its lasting influence on the field of memory rested in the concept that repetitions of the same stimuli could be encoded differently.

Since then, research in episodic memory has largely adopted encoding variability as a potential explanation for distributed practice effects; the mnemonic benefits that arise when repetitions for items studied repeatedly are spaced apart by either time or other intervening items (for reviews, see Benjamin & Tullis, 2010; Hintzman, 1974; Toppino & Gerbier, 2014). In summarizing the variables that a model of distributed practice should include, Melton (1967) offered the differences in “contextual coding” between item repetitions as a potential variable upon which distributed practice effects are contingent. In a separate work, he also suggested that this contextual independence between repetitions - that increases with spacing - allows for the item to be encoded with a larger number of cues capable of aiding memory at retrieval (Melton, 1970). These ideas are clearly consistent with both the encoding variability hypothesis proposed by Martin (1967) and the Stimulus Sampling Theory (Estes, 1955), often considered a forebear of encoding variability theories. In the case of the encoding variability hypothesis, the link lies in the suggestion that encoding variability for a memory trace does not just arise out of the individual properties of the target stimulus but also from the contextual details encoded along with the target (Melton, 1970). Stimulus sampling theory claims that memory success in repeated practice situations is contingent upon the number of contextual elements stored with each item repetition, the spacing between each repetition, and the degree to which the available contextual elements change at each repetition (Estes, 1955). Thus, in each of these accounts, the crucial elements are the independence of the encodings and the contextual cues or details stored with each item repetition.

Although the current project is only concerned with spacing as a variable to be controlled, it is necessary to consider the empirical evidence for and against encoding variability within the distributed practice literature, as testing of the encoding variability hypothesis in other domains of memory is nearly non-existent. In general, experiments seeking to determine the effects of encoding variability on distributed practice operationalize encoding variability by presenting target words in different encoding contexts at each repetition, with the remaining target words presented in the same context across repetitions and serving as the control condition against which to compare the effects of variable encoding contexts. Encoding contexts have been varied by presenting target words with different cue words (e.g., Bellezza & Young, 1989; Madigan, 1969), embedded in different sentences (e.g., Postman & Knecht, 1983), and with different orienting tasks (e.g., Bird, Nicholson, & Ringer, 1978). These manipulations are intended to emulate the hypothesized effects of the increased variability in temporal contexts associated with increases in spacing. As such, an encoding variability account of distributed practice suggests that studying words in different contexts should improve the retrieval success for words studied at short lags, attenuating the benefits of increased spacing and emulating the benefits of distributed practice.

The evidence from these manipulations has been inconsistent, to say the least, but some trends show promise for the use of encoding variability as an effective encoding strategy. In a review of practice effects, Toppino and Gerbier (2014) tabulated the results of 48 experiments implementing variable encoding context manipulations and showed that a majority of these results (~68.8%) showed significant benefits of variable encoding contexts for words studied at short lags. The results seemed to point in the opposite direction at longer lags, with memory benefitting from repeated study in the same context (~40.0%). Another review suggested that a mathematical theory of encoding variability cannot account for all empirically robust aspects of spacing functions found in the distributed practice literature (Benjamin & Tullis, 2010). Nonetheless, while it appears that the encoding variability hypothesis *alone* is not a sufficiently explanatory mechanism for distributed practice effects (see Maddox, Pyc, Kauffman, Gatewood, & Schonhoff, 2018), there is still sufficient evidence to suggest that encoding variability effects may provide insight into other memory phenomena.

Pilot studies in our lab have found that variable processing during encoding produces a significant benefit to item memory when retrieval takes place in a novel, relatively unspecified context (Diana, unpublished data). In each experiment, the cognitive context was manipulated at encoding by having participants study words paired with “Yes/No” encoding questions that required the participant to think about and process some property of the target word. All words were studied multiple times; some were presented with the same question every time (same context), others with similar questions each time (low variability), and others with different questions each time (high variability). Across four experiments, recognition performance was better for words studied in the "high variability" condition than for words studied in the "same context" condition. Unlike much of the extant literature, these studies treat encoding variability as a mnemonic strategy in its own right, rather than as an explanation for distributed practice effects. They provide solid evidence for the benefits of encoding in variable contexts, but do not offer insight into an underlying mechanism. Also missing is an understanding of how the benefits of encoding variability generalize to situations in which retrieval occurs in a previously encoded context.

One possibility is that encoding variability improves memory performance by increasing the probability of TAP. If this is the case, we might expect encoding variability to be less effective in improving performance when the retrieval context matches the encoding context. In other words, if TAP is already achieved by reinstating the encoding context at retrieval, then encoding variability should not benefit memory as much as when it is the most important factor in memory. Additionally, matching encoding processing during all study exposures to the upcoming retrieval processing should produce better memory than matching only one study exposure's processing to that at retrieval. Returning to the example above, perhaps a student who is told the exact nature of the test questions they will receive should study by practicing that question format repeatedly. An alternative possibility is that variability has benefits beyond TAP, such that at least one exposure to the test processing (achieving TAP) intermixed with variability in processing is more beneficial than repeated retrieval practice. That is, perhaps the student should study the exact nature of the test question at least once, but also study the information in a variety of other ways.

The goal of current project was to examine the effects of encoding variability on retrieval success, both in the presence and absence of TAP. Recognition memory for nouns was assessed after three study exposures in which the variability in processing at encoding from one exposure to the next served as one of the manipulations of interest. Encoding variability was operationalized via experimenter-created encoding questions. Nouns were studied in either a low, medium, or high encoding variability condition. The second manipulation of interest was whether the context at retrieval matched the context at encoding and served to determine whether potential beneficial effects of encoding variability can be attributed to TAP. The hypothesized effects are outlined below.

HYPOTHESES

1. Increased encoding variability will benefit memory for words.
2. Context match between encoding and retrieval will benefit memory for words.
3. Encoding variability will provide an additional benefit to memory even when TAP is achieved via context match.
4. Alternate hypothesis: encoding variability produces benefits only when TAP is not specifically achieved.

Cue Diagnosticity

In order to reduce the possibility of confounding effects, the specificity of the encoding contexts must be taken into consideration. Although there is abundant evidence suggesting that TAP and a general match between encoding and retrieval improve memory, both theory and empirical evidence implicate the diagnosticity of the retrieval cues as an equally important element of retrieval success (Bramão & Johansson, 2016; Nairne, 2002; Poirier et al., 2011; Goh & Lu, 2012). Cue diagnosticity - also referred to as cue specificity and closely related to the concept of cue overload (Watkins & Watkins, 1975) - refers to the degree to which retrieval cues are capable of eliciting the retrieval of a specific memory trace. If a given cue is stored with too many memory traces (cue overload) then this cue has little diagnostic value for any specific trace and is unlikely to evoke successful retrieval, even when there is a significant match between encoding and retrieval. As will be noted below, the current experiment controlled for context specificity, and therefore cue diagnosticity, by presenting every semantic encoding question with an equal number of target words and on an equal number of trials. In this way, both the external (each encoding question) and internal (processing evoked by each encoding question) context cues will be equally diagnostic of every target word.

Methods

Participants

Data were collected from 70 Virginia Tech students, recruited through the SONA Systems research participant software and from the broader Virginia Tech community. A minimum sample size (N = 30 for 96% power) was determined by a G*Power 3.1.6 analysis (Faul et al., 2007) based on the findings from a pilot study investigating encoding variability (Diana, unpublished data). Of the 70 participants, 5 were excluded from all analyses, for a final sample size of N = 65. One was excluded due to a computer error that did not allow them to complete the test phase, while four others were excluded due to poor performance (see "Analysis and Results" below for exclusion criteria). Participants recruited through SONA received course credit in one of their Psychology courses as compensation, whereas those recruited outside of the recruitment software were compensated $20 for their participation.

Design

The experiment took the form of a 3 X 2 factorial design with levels of encoding context variability (low, medium, high) and retrieval context (matching the encoding context or non-matching) as the respective variables of interest (see Figure 1). Participants were randomly assigned to one of six counterbalancing conditions to account for differences in experimental condition order.

Materials

Trial stimuli consisted of nouns, "Yes/No" questions, and response options, all presented in white font on a black background, in 24pt, 16pt, and 18pt fonts, respectively. Questions at the beginning of each block and all other texts were presented in 36pt font.

Word lists for each participant were composed of 360 concrete, four to nine letter, nouns randomly drawn from a pool of 434 nouns taken from the SUBTLEXus corpus and constrained to a SUBTLEXus frequency rating range of .02 - 292.06 per million words (*M* = 6.81) (Brysbaert & New, 2009). Words in the pool were also selected for high concreteness ratings, with a range of 4.59-5 (*M* = 4.86) (concreteness ratings range from 0-5; Brysbaert, Warriner, & Kuperman, 2014), and a minimum prevalence value of 2.00 (suggesting that approximately 98% of people are familiar with the word) based on word prevalence norms from Brysbaert, Mandera, McCormick, and Keuleers (2018).

Of the 360 words, half were randomly selected for use as study words and the other half were used as lure words. Study words were randomly assigned to either the matching or non-matching retrieval context condition. The 90-word sets were further subdivided into thirds, with 30 words randomly assigned to the low, medium, or high variability conditions. To allow for counterbalancing and to account for order effects, the 30 words in each variability condition were halved so that there were two 15 word sets in each of the six experimental conditions. The study phase was comprised of the 180 study words, presented three times each, for a total of 540 study trials. The test phase contained 360 trials, with all study words and an equal number of lure words randomly presented once.

Experimenter-created encoding questions were used to manipulate the encoding and retrieval contexts. Each of the six encoding questions required the participant to think of the item in a specific physical or relational context (e.g., “Would it hurt if this item fell on your foot?” or “Can this item be frozen in a freezer?”; see Appendix A). The questions were randomly assigned to either the matching or non-matching retrieval context condition for each participant. The words assigned to the matching condition were presented with the same question at both study and test, whereas the words assigned to the non-matching condition were presented with a novel question at test.

Procedure

The study phase was divided into seven encoding blocks, of uneven lengths, with each block representing an encoding context. In each block, participants saw a list of words and answered the same encoding question for all words in that list. Of the three encoding questions shown at study, two were presented with two separate word lists and one was presented with three separate lists, but all questions were presented for the same number of trials. Blocks, within which the encoding question did not change, varied in length (i.e., number of trials) depending on the counterbalancing scheme applied. This method of blocking trials minimized the number of times that the encoding questions changed between trials, creating a slowly changing contextual signal from list to list as opposed to the item signal which changed on each trial.

The condition orders were designed to approximate equal spacing, in terms of number of study trials intervening between encoding exposures, such that each item was repeated at a distance of zero intervening blocks (i.e. in subsequent lists). The order of items on each list was randomized so that each of the three study exposures for an item was spaced an average of 30 trials apart. This spacing control induced differences in delay between the last study exposure and the beginning of the test trials for each condition. These differences in delay were controlled by counterbalancing condition order across participants.

At the beginning of the experiment, participants were briefed on the general nature of the study and given the instructions (Appendix B) for the study phase. They were informed that their task was to study concrete nouns and were instructed to respond either “yes” or “no” to a question appearing both before each list and concurrently with each word. They were asked to do their best to respond to each word before the next trial appeared, given the quick presentation of the words. Participants were notified that some words would be repeated and to treat all words similarly. They were specifically asked not to attempt to memorize words with any specific strategy, but to simply answer the questions that appeared on screen. The idea behind this request was that participants would encode each of the words semantically and in relation to the encoding questions, rather than in an uncontrolled manner.

At the beginning of each block, participants were presented with an encoding question for 5 seconds. This same encoding question was then applied across all trials in that block. Each subsequent trial began with a 500 ms fixation cross, after which the word appeared in the center of the black screen for 1500 ms, with the encoding question directly above the word. The response option cues (“J = Yes” and “K = No”) appeared directly below the word. At the end of each block, a message on screen instructed participants to notify the experimenter that they completed the study list. The purpose of this procedure was to make each block seem like a separate encoding experience with a common temporal and mental context for each word within the block, rather than a continuous list of briefly interrupted words.

At test, participants were informed that they would be tested on their memory for each of the items presented at study. The test instructions (Appendix C) specified that participants would see a mix of study words and new words along with study questions and new questions, and that their task was to first answer “yes” or “no” to the question and then to indicate whether they studied the word before, regardless of the preceding question. They were asked to do their best on the recognition judgment and to provide their best guess if they were uncertain as to whether the word was presented at study or not. They were informed of the 10 second time limit to answer each question and asked to attempt to answer the questions as accurately as possible.

The test phase consisted of two blocks of 180 trials for a total of 360 trials. Blocks were separated by a 3-minute optional rest period to reduce potential fatigue effects. Test trials contained either a studied word or lure word paired with a familiar or novel encoding question, along with a recognition question on an ensuing screen (“Did you study this word before?”). The first trial of each block began with a 2000 ms fixation period and every subsequent trial began with a 500 ms fixation period. Following the fixation period, the test word appeared with the encoding question above it and a “Yes” or “No” response option below it, as in the study phase. Participants had up to 10 seconds to respond to the question. The requirement to respond to the question ensures that the question influences the participant’s state of mind during retrieval. Immediately after answering the encoding question, a new screen appeared displaying the same test word with the question “Did you study this word before?” above it and the response options (“Yes” or “No”) below it. Participants also had 10 seconds to respond to the recognition question before continuing to the next trial.

Analysis and Results

Item recognition was operationalized as the proportion of hits (i.e., correct responses to the test items). This value was computed for each participant in each of the six conditions. In order to establish an exclusion criterion for poor (i.e., below chance) performance, *d'* was also calculated for each condition. Participants with a *d'* score below zero in any condition were excluded from all subsequent analyses. Although *d'* was used as an exclusionary measure, proportion of hits was chosen over *d'* as the measure of recognition accuracy due to the nature of the test phase design. Because all test items and lures were presented in a randomized order and not blocked by condition, only a single false alarm rate could be computed for each participant, and thus the same false alarm rate was used to compute d' for each condition. Given this limitation, the proportion of hits was deemed a more diagnostic measure of accuracy differences between conditions.

A repeated-measures analysis of variance (ANOVA) was performed to test the proposed hypotheses. Hypotheses 1 and 2 were that encoding variability and context-match induced TAP, respectively, would each benefit item recognition. Thus, the expected results were main effects of encoding variability and context-match. Hypothesis 3, our primary hypothesis of interest, was that encoding variability would benefit performance above and beyond the benefits conveyed by TAP. Alternatively, hypothesis 4 posited that encoding variability would only benefit performance when TAP was not specifically induced.

Thus, we expected a significant interaction effect between encoding variability and context-match, such that encoding variability would benefit memory more when TAP was induced via context-matching. To specifically test for this third hypothesis, even in the absence of a significant interaction, planned comparisons were conducted via paired-samples t-test to examine mean differences between the LV-match and HV-match conditions, as well the LV-non-match and HV-non-match conditions. If the hypothesis held, then the expectation would be for higher performance in both the HV-match and HV-non-match conditions as compared with the LV-match and LV-non-match conditions. On the other hand, if high variability only benefitted performance in the non-match condition, this would lend evidence to the alternate hypothesis that encoding variability is primarily beneficial in the absence of TAP.

Condition means and marginal means are listed in Table 1. The repeated measures ANOVA and paired-samples t-tests were all performed at a significance level of .05. The ANOVA revealed significant main effects of encoding variability, $F(2, 128) = 7.00$, $p = .001$, $\eta 2 = .009$, and context-match $F(1, 64) = 4.75$, $p = .033$, $\eta 2 = .003$ (Figure 2), as predicted by hypotheses 1 and 2. The ANOVA did not reveal a significant interaction effect $F(2, 128) = 1.10$, $p = .336$, $\eta 2 = .001$, suggesting that the effects of both encoding variability and TAP on item recognition are independent of one another, contrary to what was predicted by hypothesis 3. Results from the paired-samples t-tests revealed no significant difference between the LV-match ($M = .878$) and HV-match ($M = .895$) conditions, $t(64) = -1.851$, $p = .069$, but they did reveal a significant difference between the LV-non-match ($M = .854$) and HV-non-match ($M = .891$) conditions, $t(64) = -3.625$, $p < .001$ (Figure 3). These results have two possible interpretations. The first is that there might be an interaction between encoding variability and TAP that is not being captured by our experimental design. Alternatively, they might lend support to the alternate hypothesis that encoding variability is primarily beneficial when TAP is not already being evoked.

Discussion

The purpose of the current project was to investigate the role of TAP in the beneficial memory effect of encoding variability. Specifically, we wanted to determine whether evoking TAP through encoding-retrieval context match could explain the benefits that encoding variability conveys for recognition memory. A secondary aim was to replicate the encoding variability and TAP effects in a manner that attempted to control for spacing and specificity effects. The results of the experiment demonstrate clear memory benefits of both TAP and encoding variability, as supported by prior research and our pilot experiments. However, the results with respect to the interaction between these factors are less clear. As demonstrated by the planned comparisons, there was no statistical evidence that variability benefitted memory in the matching context conditions; yet the ANOVA indicated a main effect of variability that was not modified by an interaction with TAP.

As a hypothesis, we predicted that encoding variability would provide memory benefits above and beyond those achieved by TAP alone. If this were the case, we might conclude that the encoding variability effect cannot be explained by TAP alone, and that there is another underlying mechanism driving the effect (e.g., study-phase retrieval; Thios & D’Agostino, 1976). On one hand, the simple main effect revealed in the planned t-tests of a significant difference between LV and HV conditions when the retrieval contexts do not match, but the lack of a significant difference when the contexts do match, refutes this hypothesis. If encoding variability is only improving memory when there is no match between the encoding and retrieval contexts, then TAP may be one of the primary mechanisms driving such an encoding variability effect. On the other hand, a strict interpretation of TAP would suggest that maximal performance is achieved with maximal overlap between the encoding and retrieval contexts. Therefore, encoding a single item several times and then retrieving it all in the same context (as in the LV-match condition) should allow for better memory performance than encoding an item in multiple different contexts and then retrieving it in a context that matched only one of the encoding contexts (as in the HV-match condition). In the former scenario the overlap between encoding and retrieval processing should be stronger than in the latter scenario, given the greater number of exposures to the future retrieval context. Assuming this interpretation of the TAP framework, equivalent performance in the LV-match and HV-match conditions would suggest that TAP is not the sole principle driving the encoding variability effect.

The results of our planned comparisons demonstrate no difference between the LV-match and HV-match conditions. This is consistent with a proposal that encoding variability benefits memory only when there is no match between encoding and retrieval contexts and is consistent with results from preliminary experiments in which encoding variability was beneficial to recognition in an entirely novel context. As noted above, the results are inconclusive as to whether encoding variability is driven by TAP or encoding-retrieval match in general. Although we cannot say with certainty whether the encoding variability effect is independent of TAP, we can conclude that decreasing the amount of processing match from what TAP deems ideal (i.e. from LV-match to HV-match) is not harmful when encoding variability is introduced.

Due to the high performance in this task and the small differences between conditions, we were concerned about the influence of ceiling effects on our results. As an exploratory investigation of this issue, we excluded any participants who achieved a 100% high rate in any one condition (remaining N = 38). The resulting pattern of data was comparable to that reported above (see Table 2), suggesting that the results may hold even without the presence of a ceiling effect. Of course, removing participants who perform at ceiling also removes some of the variance in the results, so this pattern should be taken with a grain of salt. Still, future studies should aim to account for such effects via one of the following methods: incorporating a longer delay between study and test, reducing the number of exposures to each item, or manipulating the properties of the items (e.g., length and concreteness) and encoding trials (e.g., shortening trial time) in order to increase difficulty and decrease overall performance.

Another factor that may have influenced our results is the inclusion of a medium-variability condition. This is particularly notable because the overall mean differences between conditions were relatively small. Given the presence of a simple main effect between the low and high variability conditions at one level of the TAP condition, but not the other, it is possible that the medium-variability condition lead to suppression of an interaction effect when included in the main ANOVA. Although the inclusion of a medium variability condition allows for better visualization of an effect gradient, future studies should aim to focus on the distinction between low and high variability conditions to more clearly establish the effect, while also controlling for confounds such as spacing and context-specificity.

Future studies should also address other possible mechanisms influencing the encoding variability effect. Our results suggest that TAP might be driving the effect, but they are not conclusive enough to suggest that other mechanisms are not also contributing. One possible alternative mechanism, that is typically implicated in the spacing literature, is study-phase retrieval (Thios & D’Agostino, 1976). Often used in an attempt to account for distributed practice effects, study-phase retrieval stipulates that spacing and lag effects should only be observed when the repetition of a studied item is recognized as having been studied before, such that the item is being retrieved at study. Subsequently, this allows for a stronger encoding of the item due to the addition of more context cues to the representation of the item in memory, the strengthening of the item memory itself, or some other unexplained mechanism. If encoding variability is an elaborative process that improves memory by adding on context cues to each repetition of an item, subsequently increasing the number of retrieval routes, study-phase retrieval would suggest that such a benefit can only be present when the repeated item is recognized as having been previously studied. Otherwise, context cues are not being added to one memorial representation of the item, but instead different and unassociated representations of the item are being formed.

While study-phase retrieval might establish a boundary condition under which encoding variability is beneficial, it is not explanatory in and of itself. It suggests that encoding variability can only work when item repetitions are recognized as previously studied but does not explain why it might lead to better memory. Additionally, recognition of a repeated item at study should be more likely to take place when the item and context are similar to the previous item presentation. In this case, a study-phase retrieval framework alone would suggest that less variability would lead to better performance at test. Therefore, to explain encoding variability effects, study-phase retrieval must be combined with some other mechanism or strategy, such as the retrieval difficulty described in Benjamin and Tullis’ (2010) reminding account of distributed practice effects. In this account, item memory benefits from study-phase retrieval more when the reminding experience is more difficult. Spacing is suggested to improve memory because of the increased difficulty of effective study-phase retrieval as a function of spacing, which leads to stronger encoding. One might expect that encoding variability works similarly, such that the more different the contexts from one exposure to the next, the harder it is to illicit study-phase retrieval, so that encoding variability is beneficial up until the point where the contexts are so dissimilar that study-phase retrieval no longer takes place.

Regardless of the mechanism involved, encoding variability is a clear example of the way in which context can be manipulated to improve episodic memory retrieval. The current experiment focused specifically on cognitive context and the cognitive processing of the target items, but it is also likely that physical context cues can be taken advantage of, so long as they are active elements of the encoding experience and have some bearing on the cognitive environment at encoding. In either case, understanding of the mechanisms that drive the encoding variability effect allows us to not just use these cues for memory improvement, but to develop theories on how context affects neural memory representations in the medial temporal lobe and associated regions. With a better understanding of these mechanisms, we might also gain insight into other memory phenomenon, such as how the brain uses context to both generalize across memories that are sufficiently similar and to distinguish between memories that are sufficiently similar. Thus, the continued study of encoding variability and context effects has wide-ranging benefits for the study of episodic memory and cognition.

Conclusion

In summary, the current project employed a behavioral manipulation to determine the role of TAP in the encoding variability effect. Building on preliminary experiments, the study controlled for spacing effects on a block level as well as for cue-diagnosticity and context-specificity. Adhering to a strict interpretation of the TAP framework, the results did not provide enough evidence to suggest that TAP drives encoding variability, nor did it provide sufficient evidence to the contrary. However, the results do suggest that encoding variability is not detrimental to memory performance, whether or not the encoding context matches the retrieval context, and even when the strictest interpretation of TAP is assumed. Additionally, the results are evidence of a consistent benefit of encoding variability when the retrieval context is novel or unknown. While the driving mechanism for encoding variability effects remains unclear, the empirical research thus far suggests that it is a practical and useful mnemonic strategy for situations in which we cannot be certain when or where we may need to retrieve a given piece of information.

References

Bellezza, F. S., & Young, D. R. (1989). Chunking of repeated events in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(5), 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*(3), 228–247. https://doi.org/10.1016/j.cogpsych.2010.05.004

Bird, C. P., Nicholson, A. J., & Ringer, S. (1978). Resistance of the Spacing Effect to Variations in Encoding. *The American Journal of Psychology*, *91*(4), 713–721. https://doi.org/10.2307/1421519

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

Bramão, I., & Johansson, M. (2016). Benefits and Costs of Context Reinstatement in Episodic Memory: An ERP Study. *Journal of Cognitive Neuroscience*, *29*(1), 52–64. https://doi.org/10.1162/jocn_a_01035

Brysbaert, M., Mandera, P., McCormick, S. F., & Keuleers, E. (2018). Word prevalence norms for 62,000 English lemmas. *Behavior Research Methods*. https://doi.org/10.3758/s13428-018-1077-9

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–990. https://doi.org/10.3758/BRM.41.4.977

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*(3), 904–911. https://doi.org/10.3758/s13428-013-0403-5

Butler, B. E., & Merikle, P. M. (1970). Uncertainty and meaningfulness in paired-Associate learning. *Journal of Verbal Learning and Verbal Behavior*, *9*(6), 634–641. https://doi.org/10.1016/S0022-5371(70)80025-1

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

Diana, R. (2017). *Context specificity vs. variability: How does it affect learning, memory, and hippocampal representations?* Unpublished Manuscript.

Diana, R. A., Yonelinas, A. P., & Ranganath, C. (2007). Imaging recollection and familiarity in the medial temporal lobe: a three-component model. *Trends in Cognitive Sciences*, *11*(9), 379–386. https://doi.org/10.1016/j.tics.2007.08.001

Goh, W. D., & Lu, S. H. X. (2012). Testing the myth of the encoding–retrieval match. *Memory & Cognition*, *40*(1), 28–39. https://doi.org/10.3758/s13421-011-0133-9

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

Lohnas, L. J., Polyn, S. M., & Kahana, M. J. (2011). Contextual Variability in Free Recall. *Journal of memory and language*, *64*(3), 249–255. https://doi.org/10.1016/j.jml.2010.11.003

Maddox, G. B., Pyc, M. A., Kauffman, Z. S., Gatewood, J. D., & Schonhoff, A. M. (2018). Examining the contributions of desirable difficulty and reminding to the spacing effect. *Memory & Cognition*. https://doi.org/10.3758/s13421-018-0843-3

Madigan, S. A. (1969). Intraserial repetition and coding processes in free recall. *Journal of Verbal Learning and Verbal Behavior*, *8*(6), 828–835. https://doi.org/10.1016/S0022-5371(69)80050-2

Martin, E. (1968). Stimulus meaningfulness and paired-associate transfer: An encoding variability hypothesis. *Psychological Review*, *75*(5), 421–441. https://doi.org/10.1037/h0026301

Melton, A. W. (1967). Repetition and Retrieval from Memory. *Science, 158*(3800), 532–532. 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 Behavior*, *9*(5), 596–606. https://doi.org/10.1016/S0022-5371(70)80107-4

Merryman, C. T., & Merryman, S. S. (1971). Stimulus encoding in the A-B’, AX-B and the A-B’r, AX-B paradigms. *Journal of Verbal Learning and Verbal Behavior*, *10*(6), 681–685. https://doi.org/10.1016/S0022-5371(71)80075-0

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning & Verbal Behavior*, *16*(5), 519-533.

Nairne, J. S. (2002). The myth of the encoding-retrieval match. *Memory*, *10*(5–6), 389–395. https://doi.org/10.1080/09658210244000216

Poirier, M., Nairne, J. S., Morin, C., Zimmermann, F. G. S., Koutmeridou, K., & Fowler, J. (2011). Memory as discrimination: A challenge to the encoding–retrieval match principle. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(1), 16. https://doi.org/10.1037/a0024956

Postman, L., & Stark, K. (1971). Encoding Variability and Transfer. *The American Journal of Psychology*, *84*(4), 461–471. https://doi.org/10.2307/1421162

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*(5), 529–536. https://doi.org/10.1016/0022-5371(76)90047-5

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

Tulving, E. (1972). Episodic and semantic memory. In *Organization of memory* (pp. xiii, 423–xiii, 423). Oxford, England: Academic Press.

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

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*(5), 352-373.

Watkins, O. C., & Watkins, M. J. (1975). Buildup of proactive inhibition as a cue-overload effect. *Journal of Experimental Psychology: Human Learning and Memory*, *1*(4), 442–452. https://doi.org/10.1037/0278-7393.1.4.442

Williams, R. F., & Underwood, B. J. (1971). Encoding variability: Tests of the Martin hypothesis. *Journal of Experimental Psychology*, *86*(2), 317. https://doi.org/10.1037/h0029998

Yonelinas, A. P. (2002). The Nature of Recollection and Familiarity: A Review of 30 Years of Research. *Journal of Memory and Language*, *46*(3), 441–517. https://doi.org/10.1006/jmla.2002.2864

Table 1.

Mean hit rate for each condition with standard deviations listed in parentheses.

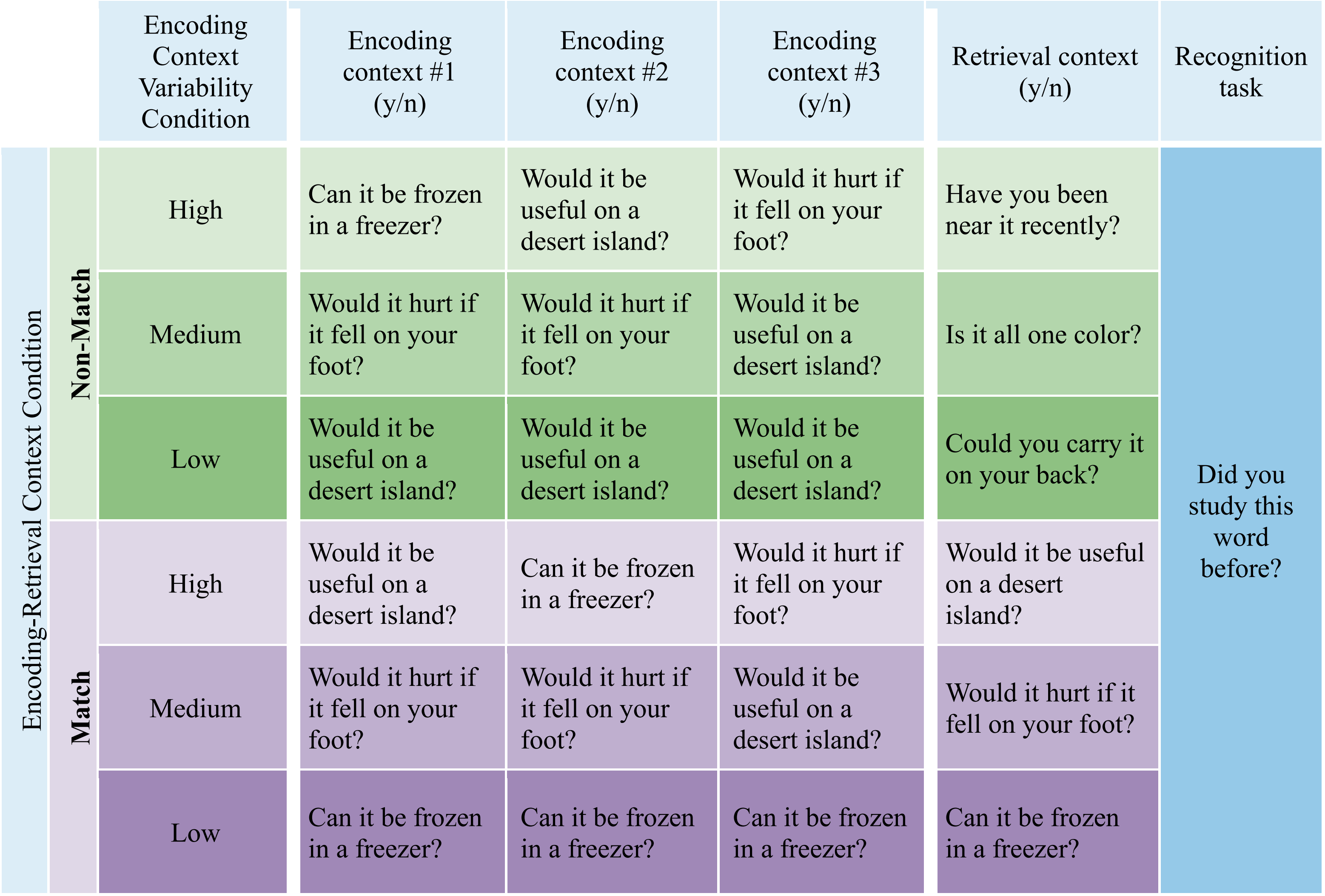
Condition	Match	Non-Match
Low	0.878 (0.112)	0.854 (0.113)
Medium	0.887 (0.125)	0.876 (0.113)
High	0.895 (0.104)	0.891 (0.110)

Table 2.

Mean hit rates for exploratory analysis excluding participants with at least one condition at ceiling. Standard deviations are listed in parentheses.

Condition	Match	Non-Match
Low	0.812 (0.109)	0.790 (0.115)
Medium	0.821 (0.136)	0.804 (0.118)
High	0.834 (0.121)	0.824 (0.117)

Figure 1. Experimental Design Matrix



		Encoding Context Variability Condition	Encoding context #1 (y/n)	Encoding context #2 (y/n)	Encoding context #3 (y/n)	Retrieval context (y/n)	Recognition task
Encoding-Retrieval Context Condition	**Non-Match**	High	Can it be frozen in a freezer?	Would it be useful on a desert island?	Would it hurt if it fell on your foot?	Have you been near it recently?	Did you study this word before?
		Medium	Would it hurt if it fell on your foot?	Would it hurt if it fell on your foot?	Would it be useful on a desert island?	Is it all one color?	
		Low	Would it be useful on a desert island?	Would it be useful on a desert island?	Would it be useful on a desert island?	Could you carry it on your back?	
	Match	High	Would it be useful on a desert island?	Can it be frozen in a freezer?	Would it hurt if it fell on your foot?	Would it be useful on a desert island?	
		Medium	Would it hurt if it fell on your foot?	Would it hurt if it fell on your foot?	Would it be useful on a desert island?	Would it hurt if it fell on your foot?	
		Low	Can it be frozen in a freezer?	Can it be frozen in a freezer?	Can it be frozen in a freezer?	Can it be frozen in a freezer?	

Figure 2. Proportion of hits by encoding variability and TAP conditions.

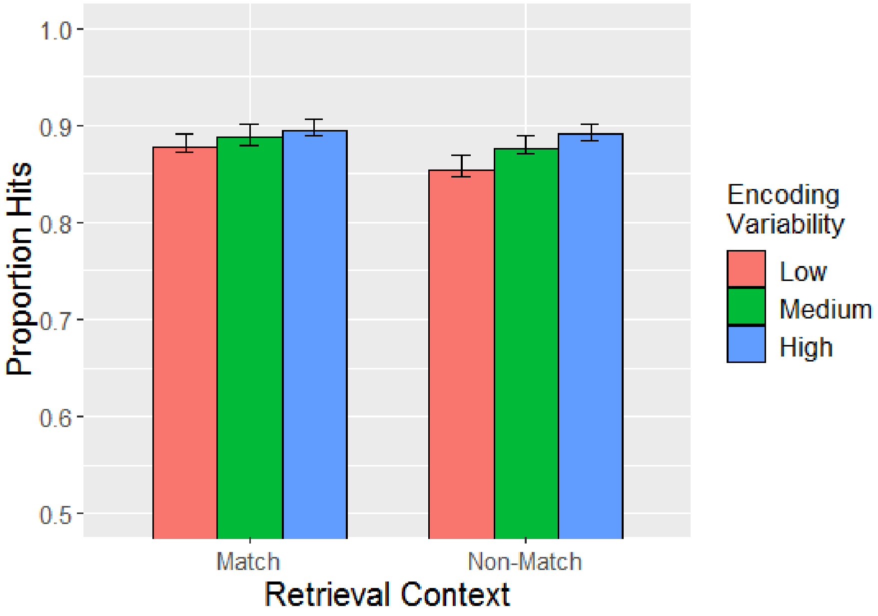
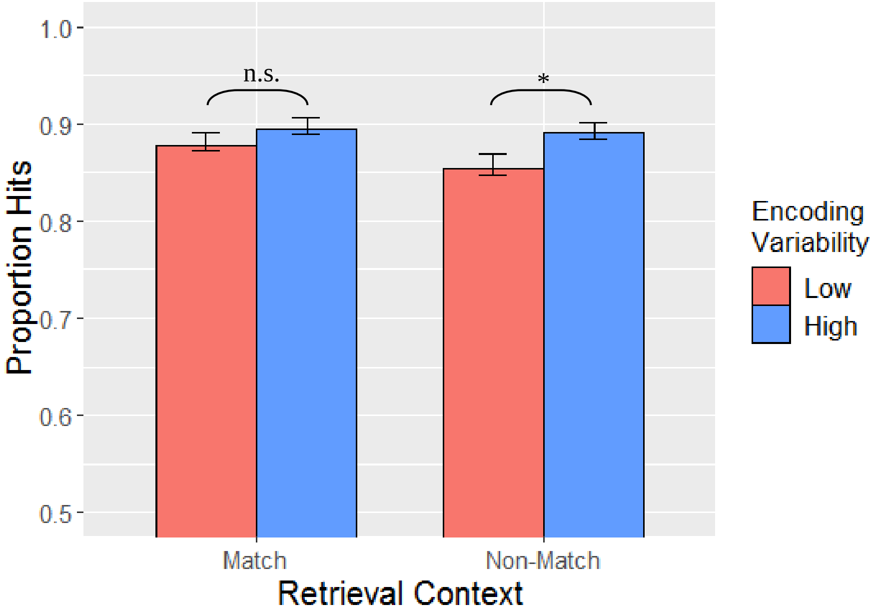


Figure 2. Proportion of hits by encoding variability and TAP conditions, excluding the medium variability condition.



Appendix A

Encoding Questions

Is this item all one color?

If you were stranded on a desert island, would this item be useful?

Can this item be frozen in a freezer?

Could you carry this item on your back?

Have you been near this item recently?

Would it hurt if this item fell on your foot?

Appendix B

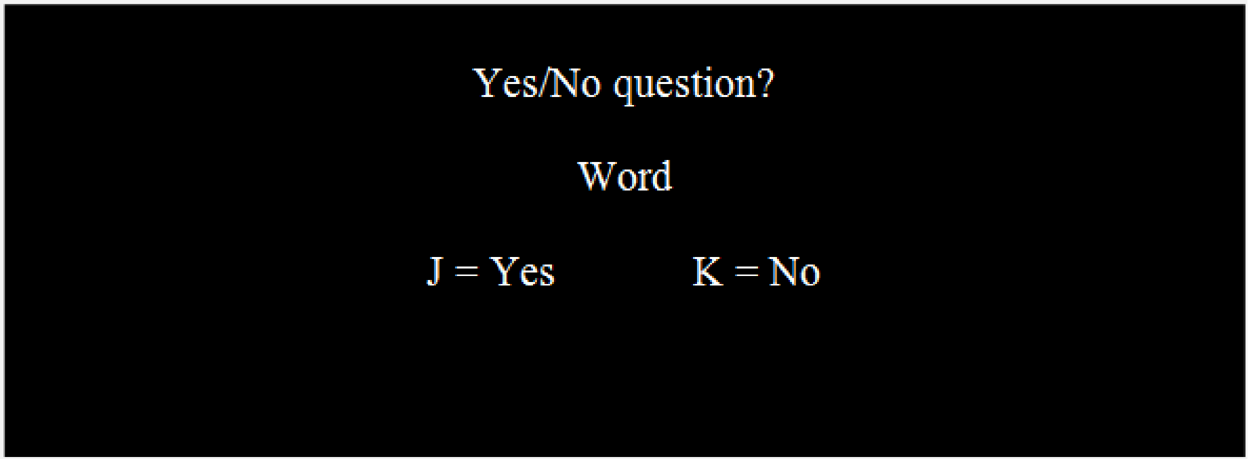
Study-Phase Instructions

In this experiment you will be studying some words for a memory test. For each word, you will be asked to make a yes/no judgment about that object. There will be multiple separate sections to the study part of the experiment and the yes/no question will be the same throughout any given section. You will see the specific question for each section at the beginning in order to have time to read it. Then each word will appear fairly quickly and we need you to try to respond “Yes” or “No” before the next word appears. The answers to the questions are entirely based on your own opinion.

Many of the words will repeat, either in the same section or a different section. Treat both repeated and new words the same way. Don’t try to use strategies to memorize the words, just answer the question.

When each study section ends, please open the door to the testing room and the experimenter will give you further instructions. Please do not attempt to continue with the next section without checking in with the experimenter.

The screen you will see on each trial looks like this:



Yes/No question?

Word

J = Yes K = No

Appendix C

Test-Phase Instructions

Now we will test you on your memory for the words that you studied in the experiment. We are interested in understanding how the yes/no questions we asked might affect your memory for the words. That means that in this part of the experiment you will again see a yes/no question presented with each word (some of the questions will be familiar and some will be new). You will have up to 10 seconds to make your yes/no judgment about the word. (The questions will be randomly mixed so you will need a bit more time to read the question with each item than you did on the study trials). Just answer the question according to your own opinion. Then you will be asked "Did you study this word before?" and be asked to answer yes or no.

The words from the study lists will be randomly mixed with new words. For each word, you should indicate whether it is a word that you studied in this experiment ("Yes" = "J") or whether it is word that you did not study in this experiment ("No" = "K"). There is no option to skip an item so just make your best guess if you aren't sure.