Journal of Memory and Language 73 (2014) 43–58

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

All varieties of encoding variability are not created equal: Separating variable processing from variable tasks

Mark J. Huff *, Glen E. Bodner

Department of Psychology, University of Calgary, Canada

ARTICLE INFO

Article history:
Received 26 September 2013
revision received 18 February 2014

Keywords:
Encoding variability
Item-specific and relational processing
Free recall
Recognition
False recognition

ABSTRACT

Whether encoding variability facilitates memory is shown to depend on whether item-specific and relational processing are both performed across study blocks, and whether study items are weakly vs. strongly related. Variable-processing groups studied a word list once using an item-specific task and once using a relational task. Variable-task groups' two different study tasks recruited the same type of processing each block. Repeated-task groups performed the same study task each block. Recall and recognition were greatest in the variable-processing group, but only with weakly related lists. A variable-processing benefit was also found when task-based processing and list-type processing were complementary (e.g., item-specific processing of a related list) rather than redundant (e.g., relational processing of a related list). That performing both item-specific and relational processing across trials, or within a trial, yields encoding-variability benefits may help reconcile decades of contradictory findings in this area.

Introduction

A classic debate in memory research is whether information that is encoded more than once is better remembered if the same or a different encoding strategy is used each time. According to the *encoding-variability (EV) hypothesis*, applying a variety of encoding strategies should be more beneficial than repeating a single encoding strategy, either because it increases the number of cues or routes that can be used to retrieve items at test (e.g., Estes, 1950), and/or because it increases the number of memory traces or the richness of a given trace (e.g., Glenberg, 1979). Although several studies have reported *EV benefits* relative to repeating an encoding task (e.g., D'Agostino & DeRemer, 1973; Greene & Stillwell, 1995; Hintzman & Stern, 1978; Hunt & Einstein, 1981; Maskarinec & Thompson, 1976; Postman & Knecht, 1983), others have shown no effect of EV (e.g., Bird, Nicholson, & Ringer, 1978; Dempster, 1987; Elmes & Bjork, 1975; Galbriath, 1975; Johnston, Coots, & Flickinger, 1972; Williams & Underwood, 1970), or even *EV costs* (e.g., Bobrow, 1970; Bower, Lesgold, & Tieman, 1969; Roediger, Sanches, & Agarwal, 2011; Young & Bellezza, 1982). The cause of these discrepant findings has never been satisfactorily determined. The present study aimed to reinvigorate research on EV by attempting to better specify some parameters that determine whether EV benefits are obtained.

Research on EV was launched by Estes' (1950) stimulus sampling theory, which characterized learning as a statistical association process between a stimulus and available contextual elements. During encoding, a study item and available encoding elements become associated, as in stimulus–response contingency learning. At test, available contextual elements can then be used to retrieve items that were sufficiently associated at encoding. Memory for an item is therefore a function of the contextual elements afforded at encoding and retrieval, and their associations with these elements. Conditions that produce variations

* Corresponding author. Address: Department of Psychology, Washington University in St. Louis, St. Louis, MO 63130, USA.

E-mail address: mhuff@wustl.edu (M.J. Huff).

http://dx.doi.org/10.1016/j.jml.2014.02.004
0749-596X/© 2014 Published by Elsevier Inc.

in stimulus elements through EV increase the number of available associations at test, thus increasing the probability of successful retrieval relative to repeated encoding conditions (see also Martin, 1968).

Other memory theories also predict EV effects. For example, Melton (1970) suggested that the encoding context establishes retrieval routes that are then utilized during test. EV conditions presumably increase the number of available retrieval routes. In a similar vein, recursive reminding theory (Hintzman, 1974, 2004) also predicts a memory advantage for variable conditions. Here, study items are mentally "tagged" during encoding. When items are repeated, participants are reminded of the initial tags but also form additional tags if the item repetition differs qualitatively from the original presentation. When repetitions do not differ, reminded tags become recursively embedded in memory, thus restricting their utility at retrieval relative to conditions with differing repetitions that produce a variety of tags. In summary, many theories predict EV benefits irrespective of whether study items are encoded with elements, retrieval routes, or tags.

The primary use of the EV hypothesis to date has been to account for the spacing effect – a memory advantage for items studied multiple times in a distributed rather than massed fashion (e.g., Balota, Duchek, & Logan, 2007; Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006; Glenberg, 1977, 1979). Spaced presentations lead participants to encode items in a variety of contexts, thereby increasing the number of possible contextual elements that can later serve as retrieval cues. In contrast, massed presentations direct participants to encode items in the same or similar contexts, resulting in a redundancy of encoded elements and the generation of fewer potential retrieval cues. In concert with the encoding specificity principle (Tulving & Thomson, 1973), which emphasizes that memory benefits from similarities between cues present at study and test–increasing the number of study contexts should also increase the probability of a context match at test. In essence, conditions that maximize the availability of retrieval cues should also increase the probability of successful remembering.

The inconsistency of EV effects is likely partly attributable to prior studies having manipulated EV in very different ways. For example, Gartman and Johnson (1972; Winograd & Geis, 1974) had participants study homographs twice that were either presented both times in the context of the same meaning (e.g., fangs-*bat*-vampire and dark-*bat*-cave) or in the context of different meanings (e.g., plate-*bat*-baseball and fangs-*bat*-vampire). Recall was greater when homographs were studied in the different meaning (i.e., variable) condition, demonstrating an EV benefit. In contrast, Postman and Knecht (1983) presented to-be-remembered items in sentences. In the repeated condition, each word was presented three times in the same sentence whereas in the variable condition, the word was presented in three different sentences. Here EV did not benefit recall.

More pertinent to our study, the effects of EV have often been examined by varying the encoding tasks used to study a set of items, rather than varying the contexts in which items are studied. Bird et al. (1978) had participants study a list of words over two study blocks. Participants either used the same or a different study task on each block. Recall was equivalent in the variable-task and repeated conditions, demonstrating that varying the task type between study blocks does not always yield an EV benefit. Using a similar design, Young and Bellezza (1982) found that recall was greater in the repeated condition–an EV cost. The authors attributed this cost to an interference process: The additional retrieval routes created through variable tasks interfered with each other at test. Young and Bellezza did not provide any direct evidence of interference, but conceded that "the memory mechanisms relating encoding variability and recall performance have yet to be clearly specified" (p. 556). Their statement holds, 30 years after.

Our study targets the possibility that EV benefits will typically occur only when *qualitatively different types of processing* are required on each study block. One well-documented processing distinction is between item-specific and relational processing (Einstein & Hunt, 1980; Hunt & Einstein, 1981; Hunt & McDaniel, 1993; see Glenberg, 1979, for other distinctions). *Item-specific processing* occurs when tasks emphasize the unique characteristics of items over relative comparisons (e.g., rating items for pleasantness or mentally visualizing individual items). *Relational processing* occurs when tasks emphasize shared characteristics and relative comparisons among items (e.g., sorting items into categories or constructing narratives with items). At test, participants who had engaged in relational processing at encoding are more likely to organize their retrieval on the basis of item relations than those who had engaged in item-specific processing.

Hunt and Einstein (1981) essentially found that combining item-specific and relational processing improved recall (see too Einstein & Hunt, 1980), though they did not interpret their findings as evidencing EV benefits, and their work has not been cited in this connection. Participants studied categorically related list items. In the relational task, study items were sorted into one of six category-labeled sets. In the item-specific task, study items were rated for pleasantness. In the *single-task groups*, items were studied once in either the item-specific or relational task. In the *repeated-task groups*, items were studied twice using the same item-specific or relational processing task each time. Their key group was the *variable-processing group*, who first studied the items using the item-specific task, and then studied the items again using the relational task (or vice versa). Free recall was highest in the variable-processing group, consistent with the EV hypothesis, and was lower and equal in the single-task and repeated-task groups. The authors argued that the relatively poor recall in the repeated-task groups was due to a *redundancy* in processing. That is, processing tasks completed twice encode less additional information than if different processing tasks were performed in each block.

In contrast, Hunt and Einstein (1981) did not find that the combination of item-specific and relational processing maximized recognition. Instead, the item-specific repeated condition was equal to the variable-processing condition which, in turn, outperformed the relational repeated condition. The equivalence between the item-specific repeated and variable-processing conditions was attributed to the recognition task's reliance on item-specific processing (see Mandler, 1980), which benefited the former condition. However, recognition levels were near ceiling in their study, making it difficult to ascertain whether test type modulated EV benefits.

Hunt and Einstein's (1981) variable-processing group completed two different study tasks, as was true of the variable-task conditions in Young and Bellezza (1982) and Bird et al. (1978), which did not produce EV benefits. Critically, however, Hunt and Einstein's variable-processing group performed *two different tasks* that required *two different types of processing* across study blocks (item-specific vs. relational). In contrast, Young and Bellezza's peg-word and narrative-construction tasks were *both relational,* and Bird et al.'s pleasantness-rating and active/passive-judgment tasks were *both item specific*. Therefore, tasks in the latter pair of studies may have resulted in redundant processing, consistent with McDaniel and Masson's (1985) surmise that EV benefits are more likely under conditions that maximize variability (and thus minimize redundancy) between conditions.

There is additional evidence for the importance of distinguishing between variable-processing and variable-task conditions. For example, Maskarinec and Thompson (1976) had participants study a list of words three times. Two of their tasks were item-specific in nature (rhyme task, syntax task) whereas the third task was relational (synonym task). Thus, variable processing was induced in this study across the three tasks in the EV condition, resulting in greater recall relative to a repeated condition.

In contrast, performing the same type of processing across blocks (e.g., two relational tasks) can even produce EV costs relative to repeating a task. Bower et al. (1969) presented participants with sets of words in each of two study blocks. Participants were instructed to integrate the items in each set into a mental image – a relational-processing task. The composition of each set remained the same or changed across blocks, analogous to repeated and variable-task conditions, respectively. An EV cost was found both in recall and recognition, suggesting that repeating the same type of processing across blocks can impair memory. One explanation for this pattern comes from Tulving (1962, 1966), who suggested that items studied in different contexts will be encoded into separate organizational structures. At retrieval, these structures have the potential to interfere with each other, resulting in a memory cost. Consistent with this claim, the contexts in which items were studied in Bower et al. varied across blocks, but the type of processing was held constant.

Reconciliation of discrepancies in the EV literature also has important applications such as educational settings. Learning assessments often depend upon a student's ability to successfully retrieve information from memory and therefore EV may provide a key method for improving retention (e.g., McDaniel & Einstein, 1989). However, the mixed state of the EV literature at present makes it difficult to recommend EV as a pedagogical framework. By attempting to specify a set of conditions that determine whether EV benefits are obtained, we hope that our research will be of interest to researchers, educators, and students alike.

Contributions of list-based processing

The effects of variable processing can also be assessed by varying whether encoding tasks are complementary or redundant with the type of processing that a study list's structure affords. Hunt and Seta (1984) suggested that increasing the number of categorically related items in a study list increases relational processing. Consistent with this claim, the materials appropriate difficulty (MAD) framework (McDaniel & Einstein, 1989) states that whether a given study task improves memory will depend on the correspondence in the processing afforded by the study task and by the study materials. *Complimentary processing* situations yield memory benefits because the processing afforded by the study task and by the study list are divergent (e.g., item-specific processing of a related list). *Redundant processing* situations do not yield memory benefits because the processing afforded by the study task and by the study list result in a high degree of overlap (e.g., relational processing of a related list). Although processing is quantitatively the same in the two cases, the qualitative processing contributions differ.

McDaniel, Einstein, and Lollis (1988) provided evidence for the interaction between encoding tasks and study materials. Participants studied either a strongly related or weakly related list using an item-specific (pleasantness rating) or relational (category sorting) task. Strongly related list items were taken from highly concrete and specified categories and were assumed to sponsor relational processing. Weakly related list items were taken from broad ad hoc categories whose relations were not readily apparent (e.g., things that are green, things made of wood) and were assumed to sponsor item-specific processing. Complimentary processing produced higher recall than redundant processing, consistent with the MAD framework.

We suggest that the benefits of complimentary (vs. redundant) processing at the study task and study list levels can be characterized as an additional type of EV because both item-specific and relational processing are performed for items. If variable processing is critical for producing EV benefits it is even possible that the source of the afforded processing (i.e., the encoding task or the study materials) may not matter. Instead, performing qualitatively different types of processing either within a single exposure to items, or across more than one exposure, may both serve to achieve the same end, namely, an EV benefit.

The present study

Hunt and Einstein's (1981) evidence for EV benefits was suggestive, but they varied both the processing type and the task type simultaneously within their paradigm. To establish the importance of variable processing, we separated these two types of EV. In Experiment 1, participants studied the same set of items across two study blocks. In the first block, item-specific processing was induced either through *mental-imagery* (MI) or *pleasantness-rating* (PR) tasks, or relational processing was induced either through *category-sorting* (CS) or *narrative-construction* (NC) tasks (after Hodge & Otani, 1996). On the second block, the same words were studied again using either the same task as the first block (*repeated-task groups*), a different task that required the same type of processing (*variable-task group*), or a different task that required the other type of processing (*variable-processing groups*; see Table 1). Relative to the repeated-task group, the use of variable processing was expected to produce an EV benefit (as in Hunt & Einstein; Maskarinec & Thompson, 1976), whereas use of variable tasks was not expected to produce an EV benefit (as in Bird et al., 1978; Bower et al., 1969; Young & Bellezza, 1982).

Across Experiment 1A and 1B memory was evaluated using two different list types known to interact with item-specific vs. relational processing. Experiment 1A used *strongly related lists* in which study items were drawn from highly concrete and specified categories. Experiment 1B used *weakly related lists* in which study items were from broad, unspecified categories. Based on prior findings (Engelkamp, Biegelmann, & McDaniel, 1998; Hodge & Otani, 1996; Hunt & Einstein, 1981; McDaniel, Einstein, et al., 1988), memory was expected to be better with under complimentary (vs. redundant) processing conditions. In this way, complimentary processing was expected to serve as a type of EV in which the processing afforded by the study task and list structure do not overlap.

Interactions between study task type and list type were expected to modulate EV benefits in the variable-processing group, such that strongly related lists might be less likely to produce EV benefits because they already foster relational processing (Hodge & Otani, 1996; Hunt & Seta, 1984). To the extent that participants spontaneously notice and make use of the relations among items to help them encode related items, the relational task block in the variable-processing group may duplicate processing participants already engage in when studying related items (Engelkamp et al., 1998; Hunt & Einstein, 1981). If so, encoding via both a relational task and an item-specific task might provide no additional benefit over encoding via two item-specific tasks. By this reasoning, variable processing may not yield an EV benefit with strongly related lists (Experiment 1A), but should with weakly related lists (Experiment 1B) where relational processing is reduced.

Memory was assessed using both an initial free recall test and a subsequent recognition test, as in Hunt and Einstein (1981) and McDaniel, Einstein, et al. (1988). As covered earlier, Hunt and Einstein found evidence for a variable-processing benefit in recall but not in recognition. Free recall benefits from the clustering of related items at test, which relational processing promotes. In contrast, recognition tests typically present items in a random order that is not conducive to benefiting from increased organization afforded by relational processing, but instead is better served by item-specific processing (McDaniel, Waddill, & Einstein, 1988). Because Hunt and Einstein's recognition rates were very high, however, it is unclear whether the equivalence between their item-specific repeated-task group and variable-processing group was due to EV being test dependent or due to a ceiling effect. Although recall and recognition tests are sensitive to different processing,

Table 1
Study conditions in Experiments 1 and 2.

Experiment/group/task(s)	Study Trial 1	Study Trial 2	*n*
Experiments 1A and 1B			
Repeated tasks			
IS–IS	MI	MI	16
IS–IS	PR	PR	16
REL–REL	CS	CS	16
REL–REL	NC	NC	16
Variable tasks			
IS–IS	MI/PR	PR/MI	16
REL–REL	CS/NC	NC/CS	16
Variable processing			
IS–REL	MI/CS	CS/MI	16
IS–REL	MI/NC	NC/MI	16
IS–REL	PR/CS	CS/PR	16
IS–REL	PR/NC	NC/PR	16
Experiments 2A and 2B			
Item-specific processing			
IS	MI	–	16
IS	PR	–	16
Relational processing			
REL	CS	–	16
REL	NC	–	16

Notes: Task order was counterbalanced across participants as denoted by a "/". Item-specific (IS) tasks were mental imagery (MI) and pleasantness ratings (PR). Relational (REL) tasks were category sorting (CS) and narrative construction (NC).

we expected that variable processing would yield benefits on both tests.

We also examined the effects of item-specific and relational processing on false memory–particularly false recognition (false recall rates were also measured but were expected to be quite low). Hunt and Einstein's (1981) recognition test presented categorically related lures. False recognition of these lures was lower in the repeated item-specific group than in either the repeated relational group or the variable-processing group. Thus, the present study also isolated the effects of repeated, variable-task, and variable processing conditions on false memory errors. Item-specific tasks tend to reduce associative false memory errors, whereas relational tasks tend to inflate them (Huff & Bodner, 2013; Hunt, Smith, & Dunlap, 2011). False recognition was therefore expected to be reduced following item-specific (vs. relational) processing in Experiments 1A and 1B.

These predicted false recognition patterns provided a critical manipulation check regarding the item-specific vs. relational nature of our study tasks (see also Hodge & Otani, 1996). In addition, whereas variable-processing groups were expected to produce the highest levels of correct memory, repeated and variable-task groups performing item-specific (vs. relational) processing at study were expected produce the lowest levels of false memory (Huff & Bodner, 2013). Thus, a dissociative pattern was expected across correct and false recognition.

In Experiment 2, interactions between processing recruited by the study task type and study list type were further evaluated (after McDaniel, Einstein, et al., 1988). Participants studied items once using one of the four Experiment 1 tasks with either strongly (Experiment 2A) or weakly (Experiment 2B) related lists. Experiment 2 was expected to replicate the complimentary processing benefits predicted in Experiment 1. In addition to providing a test of generality of this type of EV effect, the use of one study block facilitated comparisons to extant research. Based on prior studies (e.g., Hodge & Otani, 1996; Hunt & Einstein, 1981; McDaniel, Einstein et al., 1988), a complimentary processing benefit was expected in Experiment 2A for item-specific (vs. relational) tasks, and in Experiment 2B for relational (vs. item-specific) tasks.

Experiment 1A: Strongly related lists

Experiment 1A measured the effects of EV on recall and recognition using strongly related lists. Groups (see Table 1) studied a list of words twice in blocked fashion. Repeated-task groups studied each list twice using the same study task each time (e.g., MI/MI, PR/PR). Variable-task groups studied each list twice using a different item-specific or relational task each time (e.g., MI/PR, CS/NC). Variable-processing groups studied each list twice using an item-specific task once and a relational task once (e.g., MI/CS, PR/CS). Thus, in the variable-task groups only the study task varied across blocks, whereas in the variable-processing groups, both the study task and the type of processing it required varied across blocks.

Experiment 1 also examined the possibility that a type of EV benefit might occur when the type of processing sponsored by a study task does not overlap with the type of processing sponsored by a list type (McDaniel, Einstein, et al., 1988). Item-specific processing of strongly related lists should be less redundant than relational processing. If so, then performing two item-specific tasks (denoted *IS–IS group*) should produce a recall advantage relative to two relational tasks (denoted *REL–REL group*). For the same reason, recall might also be higher in the variable-processing groups (denoted *IS–REL group*) than in the REL–REL group.

False recognition was also expected to vary across the IS–IS, REL–REL, and IS–REL groups. Specifically, the IS–IS group was expected to show lower false recognition than the REL–REL group. Moreover, if a single item-specific task protects again false recognition even when a relational task is also performed, then the IS–REL group might be similar to the IS–IS group. If not, false recognition might be similar in the IS–REL and REL–REL groups.

Method

Participants

University of Calgary undergraduates participated for course credit ($N = 160$). Participants reported complete fluency in the English language. They were randomly assigned to the repeated-task group, the variable-task group, or the variable-processing group (see Table 1). Four participants were replaced for not following task instructions.

Materials

Two sets of related lists were constructed. Each set contained exemplars from four Battig and Montague (1969) categories (Set A = four-footed animals, tools, fruits, and birds; Set B = human body parts, kitchen utensils, furniture, and vegetables). The top 25 exemplars from each category were selected. The top 5 from each category served as critical lure items at test (cf. Huff, Meade, & Hutchison, 2011). The 80 study items were randomly ordered for each study trial for each participant (with the exception of the narrative construction task; see below). The set that was studied (vs. new) was counterbalanced across participants. The 120-item recognition test was composed of 40 studied list items (10 from each category, from list positions 1, 3, 5, 7, 10, 12, 14, 16, 18, and 20), 20 nonstudied critical items (5 from each category), 40 nonstudied list items (10 from each category from the same list positions), and 20 critical items from nonstudied control lists (5 from each category). The test list was randomized for each participant.

Procedure

Participants were tested individually using a computer by an experimenter who remained present throughout the procedure. They were asked to study a list of words for an unspecified memory test. They were further instructed to read each word aloud prior to completing one of the four study tasks. The study tasks were modeled after Hodge and Otani (1996), who found these tasks produced similar levels of recall and recognition. Words were presented one at a time on the screen (with the exception of the narrative construction task; see below). For the mental imagery (MI) task, participants were instructed to create a mental image of each word and to provide a vividness rating of their image using a 7-point scale (1 = least vivid, 7 = most vivid). For the pleasantness rating (PR) task, they rated each word's pleasantness on a 7-point scale (1 = least pleasant, 7 = most pleasant). For the category sorting (CS) task, they classified each word into one of the four categories listed on the screen. For the narrative construction (NC) task, five words were presented simultaneously on the screen and participants created a brief narrative silently to themselves using all five words. The five word sets were drawn at random (without replacement) from the 80 item list, but the same randomized order of these sets was then presented to all participants. Participants given the NC task also provided a vividness rating of their narrative using the MI task scale. The experimenter entered participants' verbal responses using a response box.

After the first study block participants did a 2-min arithmetic filler task. They then received the instructions for the second study task and performed the second study block. Study task order was counterbalanced across participants. After the second study block participants completed another 2-min arithmetic filler task. They then completed a 5-min free recall test with instruction to write down as many of the remembered studied words as possible on a sheet of paper. The recognition test followed immediately. Participants pressed "old" if the test item had been studied and "new" if the test item had not been studied, using labeled keys on a keyboard.

Results and discussion

The following conventions were used in the reported experiments. Results were significant at the $p < .05$ level unless otherwise indicated. Effect sizes were calculated using partial-eta-squared (η_p^2) for ANOVAs and Cohen's *d* for *t*-tests. Study block order was not included as a factor because it did not produce significant effects in preliminary analyses. Memory differences for individual tasks within a condition were rare, therefore analyses collapsed across individual task combinations. Recall and recognition means at the individual task level for all item types within each condition are presented in Appendix A. Adjusted-ratio-of-clustering (ARC) scores in recall were also calculated and these data are reported and discussed in Appendix B.

Free recall

Table 2 provides the mean proportion of correctly recalled items in the key conditions. Critical item and extra-list intrusions in recall were rare (see Appendix A) and were not analyzed.

Our first comparison examined whether correct recall differed across the repeated-task, variable-task, and variable-processing groups (.38 vs. .40 vs. .40); a one-way ANOVA revealed no differences, $F(2,157) = 1.00$, $MSE = .01$, $p = .37$. Thus, with related lists, neither variable tasks nor variable processing produced an overall EV benefit. However, inspection of Table 2 suggested that recall of related lists was higher in the IS–IS groups (repeated task and variable task) than in the REL–REL groups (repeated task and variable task), consistent with a complimentary processing benefit (McDaniel, Einstein, et al., 1988).

Two sets of analyses confirmed this pattern. First, recall was greater in the variable-task IS–IS group than in the variable-task REL–REL group (.44 vs. .36), $t(30) = 2.33$, $SEM = .02$, $d = .85$, demonstrating a complimentary processing benefit. Second, a pooled analysis revealed a recall difference across the pooled IS–IS group (repeated IS–IS and variable-task IS–IS), the pooled REL–REL group (repeated REL–REL and variable-task REL–REL), and the pooled IS–REL (i.e., variable-processing) group (.42 vs. .36 vs. .40), $F(2,157) = 4.83$, $MSE = .01$, $\eta_p^2 = .06$. Follow-up tests showed that recall was greater in the IS–IS group than the REL–REL group, $t(94) = 2.70$, $SEM = .01$, $d = .56$, and in the IS–REL group than the REL–REL group, $t(110) = 2.40$, $SEM = .02$, $d = .46$, once again showing a complimentary processing benefit. The IS–IS and IS–REL groups were equivalent, $t < 1$, demonstrating that performing item-specific processing twice on related lists was not more advantageous than performing variable processing. In sum, a form of EV benefit was found when study processing was complementary with study list type.

Correct recognition

Table 2 provides the mean recognition hit rates in the key conditions. Given the low false alarm rates to control items (Appendix A), only the hit rates were analyzed; analyses of *d'* yielded highly similar results. Hit rates were very high, as in Hunt and Einstein (1981), yet unlike in recall, recognition differed across the repeated-task, variable-task, and variable-processing groups (.91 vs. .94 vs. .94), $F(2,157) = 5.00$, $MSE = 0.004$, $\eta_p^2 = .06$. The hit rate was greater in both the variable-processing and variable-task groups than in the repeated-task group, $t(126) = 2.77$, $SEM = .01$, $d = .49$, and $t(94) = 2.03$, $SEM = .01$, $d = .42$, respectively. While near-ceiling hit rates prevent strong conclusions, these data may suggest that both variable processing and variable tasks can produce EV benefits. The variable-processing and variable-task group were equivalent, $t < 1$, but again, the near-ceiling hit rates make it unwise to overstate this null effect.

Complimentary processing between encoding tasks and study lists also benefited recognition. First, the variable-task IS–IS group outperformed the variable-task REL–REL group (.96 vs. .93), $t(30) = 2.24$, $SEM = .01$, $d = .82$. In the pooled analysis, correct recognition differed across the IS–IS, REL–REL, and IS–REL groups (.95 vs. .89 vs. .94), $F(2,157) = 12.17$, $MSE = .004$, $\eta_p^2 = .13$. As in recall, the hit rate was greater in the IS–IS group than the REL–REL group, $t(94) = 3.98$, $SEM = .01$, $d = .82$, and was also greater in the IS–REL group than REL–REL group, $t(110) = 4.26$, $SEM = .01$, $d = .81$, but was equivalent (though again near ceiling) in the IS–IS and IS–REL groups, $t < 1$. Thus, performing an item-specific task on a strongly related list produced a complementary processing type of EV benefit in recognition, as was true in recall.

Associative false recognition

Table 2 presents the mean proportions of falsely recognized critical items, which differed across the repeated-task, variable-task, and variable-processing groups (.19 vs. .25 vs. .28), $F(2,157) = 2.95$, $MSE = .04$, $p = .06$, $\eta_p^2 = .04$. False recognition was greater in the variable-processing group than in the repeated-task group, $t(126) = 2.55$, $SEM = .02$, $d = .45$, but was similar across the variable-task and repeated-task groups, $t(94) = 1.36$,

Table 2
Experiments 1A and 1B: mean (SD) proportion of recalled list items, correctly recognized list items (hits) and false alarms to critical items (CIs).

Group/tasks	Correct recall		Recognition			
	Exp. 1A	Exp. 1B	Exp. 1A		Exp. 1B	
			Hits	CIs	Hits	CIs
Repeated tasks	.38 (.10)	.36 (.07)	.91 (.08)	.19 (.17)	.93 (.07)	.11 (.10)
IS–IS	.40 (.09)	.36 (.07)	.94 (.07)	.15 (.16)	.95 (.05)	.09 (.09)
REL–REL	.36 (.12)	.36 (.08)	.88 (.08)	.24 (.17)	.90 (.08)	.13 (.10)
Variable tasks	.40 (.10)	.34 (.08)	.94 (.04)	.25 (.25)	.93 (.07)	.14 (.16)
IS–IS	.44 (.10)	.36 (.08)	.96 (.04)	.14 (.16)	.96 (.06)	.09 (.08)
REL–REL	.36 (.10)	.33 (.08)	.93 (.04)	.36 (.27)	.90 (.07)	.19 (.20)
Variable processing						
IS–REL	.40 (.08)	.39 (.08)	.94 (.05)	.28 (.22)	.95 (.05)	.13 (.13)

SEM = .04, $p = .18$, and across the variable-processing and variable-task groups, $t < 1$. The next analyses qualify the latter nonsignificant effects.

Consistent with prior demonstrations that item-specific processing reduces associative false recognition (e.g., Huff & Bodner, 2013), false recognition occurred less than half as often in the variable-task IS–IS (vs. REL–REL) group (.14 vs. .36), $t(30) = 2.82$, $SEM = .05$, $d = 1.03$. The pooled IS–IS, REL–REL, and IS–REL groups also differed in false recognition of critical items (.14 vs. .28 vs. .28), $F(2,157) = 7.58$, $MSE = .04$ $\eta_p^2 = .09$. False recognition was lower in the IS–IS group than in either the REL–REL group (.14 vs. .28), $t(94) = 3.47$, $SEM = .04$, $d = .72$, or the IS–REL group (.14 vs. .28), $t(110) = 3.66$, $SEM = .04$, $d = .70$. This pattern replicates prior findings that relational processing increases associative false recognition (e.g., Huff & Bodner, 2013; McCabe, Presmanes, Robertson, & Smith, 2004). False recognition was equivalent across the IS–REL and REL–REL groups (.28 vs. .28), $t < 1$; thus Experiment 1A is the first to show that item-specific processing does not reduce associative false recognition if a relational task was also performed on the same items at study. Finally, and importantly, false recognition aligned with the expected pattern (i.e., relational > item-specific), supporting our claim that our two types of study tasks produced the expected type of processing (see also Hodge & Otani, 1996).

Summary

With strongly related lists, neither variable processing nor variable tasks produced an overall EV benefit on recall relative to repeated tasks. This was expected for variable processing, given that performing a relational task is redundant with the processing induced by related materials (McDaniel, Einstein, et al., 1988). In contrast, variable processing and variable tasks both produced EV benefits in recognition, though we acknowledge that these small effects may reflect a restricted range due to ceiling hit rates.

A key finding in both tasks was a complimentary processing benefit (McDaniel, Einstein, et al., 1988), reflecting a memory advantage when item-specific processing was performed on strongly related lists. These results provide evidence of an additional type of EV benefit: Variable processing induced by a non-redundant combination of study task and materials. This effect suggests that performing variable processing, rather than performing multiple study trials per se, may be a key to producing EV benefits, a possibility we further evaluate in Experiment 2.

Experiment 1B: Weakly related lists

Experiment 1B was a replication of Experiment 1A using lists consisting of items from four broad, unspecified categories (Hunt & Einstein, 1981; Van Overschelde, Rawson, & Dunlosky, 2004). The words on these lists were ostensibly unrelated unless the participant was provided with the category labels, and were certainly less obviously related than the Experiment 1A lists. Relational processing tasks should be less redundant with weakly related lists than with the related lists used in Experiment 1A, and were therefore expected to be more effective here. If so, then variable processing should generate an overall EV benefit relative to variable tasks.

With weakly related lists, relational processing should now benefit recall and recognition to the extent that these lists sponsor less relational processing than strongly related lists. Relational study tasks could be particularly advantageous for later recall of weakly related items, by providing an organizational structure for retrieval not provided by item-specific tasks. Alternatively, participants might continue to seek relations even among ostensibly unrelated list items. For instance, Tulving (1962, 1968) suggested that participants organize information at study based on their idiosyncratic perceptions of order. If participants attempt to organize unrelated items then the complimentary processing benefits with strongly related lists in Experiment 1A might disappear.

Method

Participants

Additional University of Calgary undergraduates participated for course credit ($N = 160$). They reported complete fluency in the English language. They were randomly assigned to the same groups tested in Experiment 1A. Two participants were replaced for not following task instructions. Experiments 1A and 1B were conducted during separate semesters.

Materials and procedure

The weakly related lists were modeled after those used in Hunt and Einstein (1981) and Van Overschelde et al. (2004). Van Overschelde et al.'s norms did not contain enough items. Therefore, two new sets of weakly related lists were created, each containing four broad categories with 25 exemplars (Set A = liquids, things women wear, things in a kitchen, and things that are green; Set B = things that make noise, things made of wood, things that are soft, and things that are black). These exemplars were chosen based on ratings from 7 additional participants who rated how typical each item was for its category. The top five most typical exemplars were designated as critical lures at test to match Experiment 1A. The Experiment 1A procedure was used.

Results and discussion

Free recall

Table 2 provides the correct recall means. Unlike in Experiment 1A, a one-way ANOVA revealed a difference in recall across the repeated-task, variable-task, and variable-processing groups (.36 vs. .34 vs. .39), $F(2,157) = 3.94$, $MSE = .01$, $\eta_p^2 = .05$. Critically, the variable-processing group outperformed both the variable-task group, $t(94) = 2.37$, $SEM = .02$, $d = .49$, and the repeated-task group, $t(126) = 2.21$, $SEM = .02$, $d = .39$; the repeated-task and variable-task groups did not differ, $t < 1$. This pattern provides important new evidence that variable processing can produce an EV benefit.

We next examined whether a complimentary processing benefit occurred when relational processing was performed on weakly related lists. Contrary to this pattern, recall was equivalent in the variable-task IS–IS and variable-task REL–REL groups (.36 vs. .33), $t < 1$. The pooled analyses as reported in Experiment 1A were also conducted. Recall differed across the pooled IS–IS, REL–REL, and IS–REL (i.e., variable-processing) groups (.35 vs. .35 vs. .39), $F(2,157) = 4.14$, $MSE = .01$, $\eta_p^2 = .05$. Confirming the EV benefit reported above, recall was greater in the IS–REL group than in either the IS–IS group, $t(110) = 2.32$, $SEM = .02$, $d = .44$, or the REL–REL group, $t(110) = 2.44$, $SEM = .02$, $d = .47$. The IS–IS and REL–REL groups were equivalent, $t < 1$.

In sum, the variable-processing group showed an EV benefit relative to the repeated-task and variable-task groups. Second, a complimentary processing benefit did not emerge with weakly related lists (cf. Experiment 1A; McDaniel, Einstein, et al., 1988).

Correct recognition

Table 3 reports the correct recognition means. Hit rates were again very high. Correct recognition differed across the repeated-task, variable-task and variable-processing groups (.93 vs. .93 vs. .95), $F(2,157) = 3.66$, $MSE = .004$, $\eta_p^2 = .05$. Convergent with recall, recognition was greater in the variable-processing group than in either the repeated-task group, $t(126) = 2.66$, $SEM = .01$, $d = .47$, or the variable-task group, $t(94) = 2.09$, $SEM = .01$, $d = .43$. The repeated-task and variable-task groups were equal, $t < 1$. Thus, variable processing produced EV benefits in recall and recognition alike.

A complimentary processing benefit did not occur in recognition, as was true in recall. Contrary to this possibility, correct recognition was *greater* in the variable-task IS–IS group than in the variable-task REL–REL group (.96 vs. .90), $t(30) = 2.82$, $SEM = .02$, $d = 1.03$. This result suggests that weakly related lists may sponsor relational (than item-specific) processing, either explicitly via subjective list organization during encoding (e.g., Tulving, 1962), or perhaps via automatic spreading activation (Anderson, 1983; Collins & Loftus, 1975), as discussed later. In addition, given that recognition is sensitive to item-specific processing (Hunt & McDaniel, 1993; McDaniel, Waddill, et al., 1988), the variable-task IS–IS group may have benefited from this sensitivity (i.e., transfer-appropriate processing; Morris, Bransford, & Franks, 1977).

In the pooled analyses, the IS–IS, REL–REL, and IS–REL groups differed (.95 vs. .90 vs. .95), $F(2,157) = 14.35$, $MSE = .003$, $\eta_p^2 = .16$. Recognition was greater in the IS–REL group than REL–REL group, $t(110) = 4.71$, $SEM = .01$, $d = .90$, and was greater in the IS–IS group than the REL–REL group, $t(94) = 4.02$, $SEM = .01$, $d = .83$, but recognition was once again similar but at ceiling in the IS–REL and IS–IS groups, $t < 1$, rendering this null difference moot.

Associative false recognition

Table A4 provides the mean false recognition rates for critical items. As expected, weakly related lists generated less false recognition than strongly related lists, hence false recognition was equivalent across the repeated-task, variable-task, and variable-processing groups (.11 vs. .14 vs. .13), $F(2,157) = 1.18$, $MSE = .02$, $p = .31$. False recognition was somewhat lower in the variable-task IS–IS group than in the variable-task REL–REL group (.09 vs. .19), but this difference was marginal, $t(30) = 1.75$, $SEM = .04$, $p = .09$, $d = .64$. Pooled analyses were also conducted. False recognition differed across the IS–IS, REL–REL, and IS–REL groups (.09 vs. .15 vs. .13), $F(2,157) = 3.14$, $MSE = .02$, $\eta_p^2 = .04$. False recognition was lower in the IS–IS group than in either the REL–REL group, $t(94) = 2.43$, $SEM = .02$, $d = .50$, or the IS–REL group, $t(110) = 2.17$, $SEM = .02$, $d = .41$. The REL–REL and IS–REL groups were equivalent, $t < 1$, replicating Experiment 1A. Completing a single relational task increased false recognition such that performing item-specific processing failed to protect participants from this memory illusion at test.

Summary

With weakly related lists, variable processing produced EV benefits – in both recall and recognition – relative to both variable tasks and repeated tasks. These EV benefits were also found when the variable-processing (i.e., IS–REL) group was compared to the pooled IS–IS and REL–REL groups. As discussed below, the pattern across Experiments 1A and 1B shows that study list type modulates a subset of EV effects. In contrast to Experiment 1A, a complimentary processing benefit did not occur with weakly related lists. Recall was equivalent following relational and item-specific tasks, and recognition was actually greater following item-specific tasks. Thus, weakly related lists may continue to sponsor relational processing, making relational processing tasks redundant. This possibility is discussed further in the General Discussion.

Table 3
Experiments 2A and 2B: mean (SD) proportion of recalled list items, correctly recognized list items (hits) and false alarms to critical items (CIs).

Group/task	Correct recall		Recognition			
	Exp. 1A	Exp. 1B	Exp. 1A		Exp. 1B	
			Hits	CIs	Hits	CIs
Item-specific processing						
Mental imagery	.37 (.10)	.25 (.08)	.93 (.05)	.15 (.10)	.89 (.08)	.12 (.09)
Pleasantness rating	.29 (.09)	.24 (.08)	.87 (.14)	.24 (.16)	.85 (.14)	.12 (.10)
Relational processing						
Category sorting	.28 (.06)	.31 (.06)	.83 (.10)	.28 (.18)	.88 (.08)	.24 (.22)
Narrative construction	.27 (.09)	.22 (.12)	.82 (.11)	.26 (.17)	.78 (.16)	.31 (.23)

Experiment 2A: Strongly related lists

Experiment 2 further evaluated the benefits of complimentary processing across study task and study list, which we suggest is a type of EV. Although Experiment 1A found a benefit when an item-specific task was combined with a strongly related list, the reciprocal pattern was not found in Experiment 1B when a relational task was combined with a weakly related list. In Experiment 1, participants studied lists over two blocks, making it unclear whether the complimentary benefits in Experiment 1A were solely attributable to the type of processing completed or were also influenced by the occurrence of a second study block. In Experiment 2 we therefore used a single study block to evaluate the interaction between task and list type. Experiment 2A was similar to Experiment 1A but with only one study block. With strongly related lists, recall and correct recognition were expected to be greater following item-specific (vs. relational) tasks, whereas false recognition was expected to be lower following item-specific (vs. relational) tasks (i.e., a mirror-effect pattern).

Method

Additional University of Calgary undergraduates participated for course credit ($N = 64$). They reported complete fluency in the English language. They were randomly assigned to one of the four study tasks (see Table 1). The experiment was identical to Experiment 1A except participants studied the list of words only once using one of the four tasks.

Results and discussion

Free recall

Table 3 provides the correct recall means. We first examined whether recall could be pooled across the two item-specific tasks, and across the two relational tasks, toward an overall analysis of processing type. Recall was equivalent in the CS and NC relational tasks (.28 vs. .27), $t < 1$, but was greater in the MI (vs. PR) item-specific task (.37 vs. .29), $t(30) = 2.02$, $SEM = .03$, $p = .05$, $d = .74$, precluding the pooling of tasks within each processing type. Instead, a significant one-way ANOVA on recall across the four tasks justified comparing the remaining task pairs, $F(3, 60) = 3.61$, $MSE = .01$, $\eta_p^2 = .15$. Evidencing a complimentary processing benefit, recall was greater in the IS MI task than in either the REL CS task (.37 vs. .28), $t(30) = 2.58$, $SEM = .03$, $d = .94$, or the REL NC task (.37 vs. .27), $t(30) = 2.56$, $SEM = .03$, $d = .93$, consistent with Experiment 1A. However, the IS PR task did not differ from either the REL CS task (.29 vs. .28), $t < 1$, or the REL NC task (.29 vs. .27), $t < 1$. Thus a complimentary processing benefit occurred for the MI task but not for the PR task.

Correct recognition

For consistency with our recall analyses, the correct recognition means (Table 3) were not pooled across tasks within a given processing type. Recognition was equivalent between the REL CS and REL NC tasks (.83 vs. .82), $t < 1$, but marginally greater between the IS MI than the IS PR tasks (.93 vs. .87), $t(30) = 1.77$, $SEM = .03$, $p = .09$, $d = .65$. Thus means were not pooled across processing type. A significant one-way ANOVA justified the remaining pairwise comparisons, $F(3,60) = 3.96$, $MSE = .01$, $\eta_p^2 = .17$. As was true of recall, a complimentary processing benefit occurred only for the MI item-specific task. Specifically, recognition after the IS MI task was greater than in either the REL CS task (.93 vs. .83), $t(30) = 3.81$, $SEM = .03$, $d = 1.39$, or the REL NC task (.93 vs. .82), $t(30) = 3.80$, $SEM = .03$, $d = 1.39$. In contrast, recognition after the IS PR task was similar to both the REL CS task (.87 vs. .83), $t < 1$, and the REL NC task (.87 vs. .82), $t(30) = 1.10$, $SEM = .03$, $p = .28$.

Associative false recognition

Mirroring recall and correct recognition, false recognition (Table 3) was lower in the MI than PR item-specific task (.15 vs. .24), $t(30) = 2.05$, $SEM = .04$, $p = .05$, $d = .75$, but was similar in the CS and NC relational tasks (.28 vs. .26), $t < 1$. Further mirroring recall and correct recognition, false recognition was lower after the IS MI task than after the REL CS task (.15 vs. .28), $t(30) = 2.75$, $SEM = .05$, $d = 1.00$, or the REL NC task (.15 vs. .26), $t(30) = 2.26$, $SEM = .05$, $d = .83$. In contrast, false recognition after the IS PR task was similar to both the REL CS task (.24 vs. .28) and the REL NC task (.24 vs. .28), $ts < 1$. Though consistent across recall, correct recognition, and false recognition, the dissociation between the two item-specific tasks was unexpected given their similarity across two study blocks in Experiment 1A.

Summary

Experiment 2A provided mixed evidence for a complimentary processing benefit when an item-specific task was used to study a strongly related list. This form of EV benefit occurred with the MI task but not with the PR task. The MI task may have been more effective at sponsoring item-specific processing. Consistent with this possibility, the MI task reduced false recognition relative to the two relational tasks, but the PR task did not. The latter result is perhaps not surprising given that this task sometimes decreases false recognition (Huff & Bodner, 2013) and sometimes increases it (Toglia, Neushatz, & Goodwin, 1999), depending on the comparison group used (see Hunt et al., 2011). On the other hand, the IS–IS group produced both a complimentary processing benefit and reduced false recognition relative to the REL–REL group. The pattern across Experiments 1A and 2A suggests that MI and PR both sponsor item-specific processing, but the PR task may need to be performed twice to match the level obtained using the MI task.

Experiment 2B: Weakly related lists

Experiment 2B replicated Experiment 2A using weakly related lists. The key question was whether relational (vs. item-specific) tasks would produce a complimentary processing benefit with weakly related lists (cf. Experiment 1B).

Method

Additional University of Calgary undergraduates participated for course credit ($N = 64$), and were randomly assigned to one of the four study tasks (see Table 1). Two participants were replaced for not following test instructions. The experiment was identical to Experiment 1B except participants studied the list of words only once using one of the four tasks. Experiments 2A and 2B were conducted in separate semesters.

Results and discussion

Free recall

Table 3 reports the correct recall means. We first examined whether the two item-specific tasks, and the two relational tasks, could each be pooled. Recall was greater in the CS task than the NC task (.31 vs. .22), $t(30) = 2.69$, $SEM = .02$, $d = .98$, but was equivalent between the MI and PR tasks (.25 vs. .24), $t < 1$. Tasks were therefore not pooled by processing type as in Experiment 2A. Instead, given the significant one-way ANOVA, $F(3,60) = 3.12$, $MSE = .01$, $\eta_p^2 = .14$, the remaining pairs of tasks were compared. As in Experiment 2A, a complimentary processing benefit occurred for one of the two relational tasks. Recall was greater after the REL CS task than after either the IS MI task (.31 vs. .25), $t(30) = 2.32$, $SEM = .02$, $d = .85$, or the IS PR task (.31 vs. .24), $t(30) = 2.92$, $SEM = .02$, $d = 1.07$. However, the REL NC task did not differ from either the IS MI task (.22 vs. .25), $t < 1$, or the IS PR task (.22 vs. .24), $t < 1$. Thus, only the CS task produced a complimentary processing benefit.

Correct recognition

Correct recognition (see Table 3) was greater after the CS than the NC task (.88 vs. .78), $t(30) = 2.43$, $SEM = .03$, $d = .89$, and equivalent after the IS and PR tasks (.89 vs. .85), $t < 1$. The remaining task pairs were compared given a significant one-way ANOVA, $F(3,60) = 3.01$, $MSE = .01$, $\eta_p^2 = .13$. The only significant difference was between the IS MI and REL NC tasks (.89 vs. .78), $t(30) = 2.56$, $SEM = .03$, $d = .93$ (remaining $ts < 1.43$, $ps > .16$). Thus the relational tasks did not display a complimentary processing benefit in recognition, unlike in recall.

Associative false recognition

False recognition (see Table 3) was equivalent following the MI and PR item-specific tasks (.12 vs. .12), $t < 1$, and following the CS and NC relational tasks (.24 vs. .31), $t < 1$. False recognition was also lower after the IS MI task relative to both the REL CS task (.12 vs. .24), $t(30) = 2.10$, $SEM = .03$, $d = .77$, and the REL NC task (.12 vs. .31), $t(30) = 2.80$, $SEM = .03$, $d = 1.02$. False recognition was marginally lower after the IS PR task relative to the REL CS task (.12 vs. .24), $t(30) = 1.72$, $SEM = .04$, $p = .10$, $d = .63$, but significantly lower than after the REL NC task (.12 vs. .31), $t(30) = 2.92$, $SEM = .03$, $d = 1.07$, consistent with the processing expectations of the tasks.

Summary

Experiment 2B provided little evidence of a complimentary processing benefit when a relational task was performed on weakly related items. Here, this evidence was restricted to recall and only for the CS task. In recognition, the REL NC task resulted in *worse* recognition than item-specific tasks, suggesting the NC task may sponsor less relational processing than the CS task. On the other hand, false recognition was equivalent following the NC and CS tasks, suggesting the two sponsored equivalent relational processing. Interestingly, the repeated-task NC–NC group in Experiment 1B produced the same pattern: Lower correct recognition but equivalent false recognition to the repeated-task CS–CS group.

General discussion

Our study provides several new insights into the ingredients that can produce EV benefits in recall and recognition. By implementing Hunt and Einstein's (1981) distinction between item-specific and relational processing, the effects of two types of EV were separated and evaluated across two study trials in Experiment 1. With strongly related lists (Experiment 1A), neither variable-task nor variable-processing groups produced an overall EV benefit in recall relative to a repeated-task group, but both groups showed EV benefits in recognition. With weakly related lists (Experiment 1B), in contrast, EV benefits emerged in both recall and recognition, but importantly, only in the variable-processing groups.

A third type of EV was also identified based on complimentary processing–that memory can benefit when the type of processing required by a study task does not overlap much with the type of processing normally sponsored by a given type of list (McDaniel, Einstein, et al., 1988). Based on the MAD framework (McDaniel & Einstein, 1989), EV benefits might be expected when related lists receive item-specific processing (Experiments 1A and 2A), and when unrelated lists receive relational processing (Experiments 1B and 2B). We often (but not always) obtained such complimentary processing benefits. With strongly related lists, item-specific tasks produced this benefit relative to relational tasks in Experiment 1A, but only the MI task showed this benefit in Experiment 2A. With weakly related lists, the pattern was less consistent. Relational tasks did not produce this benefit relative to item-specific tasks in Experiment 1B, and only the relational CS task showed this benefit and only in recall in Experiment 2B.

We also examined the influence of item-specific and relational tasks on associative false recognition. As others have found (e.g., Hunt & Einstein, 1981), false recognition was lower following item-specific (vs. relational) tasks. Further, completing item-specific (vs. relational) tasks in both blocks maximally reduced false recognition, irrespective of whether the same (repeated-task) or different (variable-task) item-specific task was performed. These patterns provide converging evidence that our item-specific and relational tasks generally sponsored their expected type of processing. Of note, false recognition was equivalent between the IS–REL (i.e., variable-processing) and REL–REL conditions across list types. Interestingly, completing an item-specific task did not protect memory from associative memory illusions if a relational task was also performed on the same set of items.

On the hunt for EV BENEFITS

By separating two types of EV—namely variable processing and variable tasks—our study helps provide some resolution of the long-inconsistent EV literature. Specifically, we propose that performing both item-specific *and* relational processing at study, either across trials or within trials by crossing processing type with list type (i.e., complimentary processing), may generally foster EV benefits. Prior studies have relied on variable-task conditions to test the EV hypothesis (e.g., Greene & Stillwell, 1995; Postman & Knecht, 1983; Young & Bellezza, 1982), but prior evidence that variable processing improves memory (Einstein & Hunt, 1980; Hunt & Einstein, 1981) was overlooked. The present experiments are the first to compare variable processing to both variable task and repeated task conditions.

Three potential boundary conditions for EV benefits were also identified. First, variable tasks generally did not produce EV benefits (cf. recognition in Experiment 1A). Second, variable processing EV benefits were limited to weakly related study lists. Third, the complementary processing type of EV benefit was limited to strongly related lists. That reliable EV benefits are not ubiquitous may help resolve the discrepant patterns reported in the EV literature.

Accounts of EV effects

Our study helps inform accounts of EV effects. As reviewed in our introduction, EV has long been postulated to occur when multiple contextual elements become associated with study items (Estes, 1950; Hintzman, 1974; Melton, 1970). Such accounts do not acknowledge a role for processing type in modulating the strength of these associations. The sensitivity of EV effects to processing type suggests that EV benefits are not due to the sheer number of elements encoded, but critically depend on qualitative aspects of these associations. Performing both item-specific and relational processing at study fosters different types of associations between elements and study items that confer different potential memory benefits at test.

By acknowledging the possibility that item-specific and relational processing lay down qualitatively different associations between encoded elements and study items, the notion of variable processing brings greater specificity to the traditional EV view. We have adopted a qualitative processing-based view of these associations rather than assuming a quantitative difference in the number of encoded elements across item repetitions (cf. Johnston & Uhl, 1976). As such, variable processing likely operates to bias the associations formed with specific contextual elements. In contrast, variable tasks more closely approximate the traditional EV view, however, this condition typically failed to produce EV benefits. Thus, biasing the types of associations formed between elements and studied items through variable processing may drive EV benefits.

A remaining issue is how EV modifies associations between study items and encoded elements. Two accounts may provide insight into this association process. Hintzman's (2004) recursive reminding hypothesis suggests that reminding of a study item's previous occurrence is required for the encoding of novel elements. This hypothesis does not specify how elements become associated to study items per se, only that subsequent presentations of items must remind a participant of the initial presentation. Separately, component-levels theory (Glenberg, 1979) suggests that several types of elements are encoded over multiple presentations and that these elements interact to enhance memory. Like component levels, the item-specific/relational processing approach to EV highlights the importance of qualitative differences across encodings.

The qualitative aspect of variable processing in producing EV benefits could also have implications for accounts of other memory phenomena such as the spacing effect (Glenberg, 1977, 1979) and the retrieval-practice effect (Roediger & Karpicke, 2006). Researchers have invoked the EV hypothesis to account for memory benefits such as these. If EV contributes to these effects then it should be possible to show that variable processing is induced across study and/or test trials in these paradigms as well, providing an important avenue for future research.

Complimentary processing benefits: EV within a single study trial?

Our experiments highlight the role of study list structure in shaping whether a given type of processing will benefit memory. As reviewed above, the notion that different types of study list sponsor different types of processing is not new (e.g., Hunt & Seta, 1984; McDaniel, Einstein, et al., 1988). In general, the effectiveness of an item-specific or relational study task is inversely related to the type of processing fostered by the study materials has also been well documented (Hunt & Einstein, 1981; McDaniel, Einstein, et al., 1988). We extended these classic demonstrations by showing interactive effects between task- and list-sponsored processing across study blocks, thus linking these effects with the EV literature.

Consistent with the MAD framework, we found evidence that strongly related lists sponsor relational processing, and weakly related lists sponsor less relational processing (rather than sponsoring item-specific processing). Moreover, we obtained complimentary processing benefits for MI and PR item-specific tasks paired with strongly related lists in Experiment 1A, but only for the MI task in Experiment 2A. In contrast, these benefits were less consistently observed with weakly related word lists; a complimentary processing benefit was found in Experiment 2B, but only for the CS task and only in recall. Thus, the predicted benefits of processing overlap from the MAD framework were not always obtained.

Although it remains to be determined why complimentary processing benefits were sometimes absent, one possibility is that some of our tasks may have been more process pure than others. Memory tests are not generally process pure (e.g., Jacoby, 1991), and the same is likely true of encoding tasks as well. For instance, the type of relational processing sponsored by our CS and NC tasks likely differed. The CS task emphasizes the semantic relations among study items, whereas the NC task likely emphasizes temporal associations essential for story construction. Although researchers including ourselves have grouped tasks into neat processing typed clusters (e.g., Hodge & Otani, 1996), unique subtypes of item-specific and relational tasks may exist. The item-specific/relational processing framework has been silent to date regarding this possibility. Whether there are different subtypes of item-specific processing (e.g., perceptual vs. contextual) or relational processing (e.g., semantic vs. temporal) is therefore an important avenue for future research. Regardless, such differences would not compromise our finding that variable processing can produce EV benefits.

On a related note, the type of processing sponsored by a given list type may also not be process pure. For example, the absence of complimentary processing benefits with weakly related lists suggest that such lists still sponsor relational processing that would render relational tasks redundant. In addition, associative false recognition still occurred with weakly related lists, suggesting that the associative structure of these lists was sufficient to produce memory errors. Consistent with Tulving's (1962) argument that participants naturally organize study items, even completely unrelated lists may sponsor relational processing. Thus, relational processing may represent the default type of processing participants utilize when studying lists of items. If so, it may prove difficult to construct lists that sponsor only item-specific processing, and hence that produce a complimentary processing benefit with relational tasks.

Other evidence suggests that participants automatically establish relations between items that are, at least on the surface, unrelated. For example, mediated priming refers to facilitation in responding to a target word (e.g., *box*) that is preceded by prime (e.g., *beach*) related only indirectly through a nonpresented mediator (e.g., *sand*), relative to prime-target pairs that do not share a mediator (Balota & Lorch, 1986; Chwilla & Kolk, 2002). This facilitation also occurs when study items that are indirectly related (and therefore appear unrelated) produce false recognition of an indirectly related nonpresented item (Huff, Coane, Hutchison, Grasser, & Blais, 2012). In sum, even with unrelated lists, participants may perform relational processing either via intentional subjective organization (e.g., Tulving, 1962) or a more automatic process (e.g., Anderson, 1983).

A few additional future directions

Above we suggested several areas for follow-up research. Here we note a few limitations of our experiments that should also be remedied.

To begin, we obtained near-ceiling recognition rates, precluding strong conclusions about whether variable processing produces a larger EV benefit than variable tasks. Of course, this is not an unusual outcome when a recognition test follows initial recall (e.g., Burns, Burns, & Hwang, 2011; Hunt & Einstein, 1981; McDaniel, Einstein, et al., 1988). Anticipating this, we took steps to keep recognition off ceiling here, including using long study lists and filler tasks, to little avail. Given that initial recall testing can inflate subsequent correct recognition (e.g., Huff et al., 2012; Roediger & McDermott, 1995), eliminating the initial recall test would reduce hit rates allowing clearer conclusions to be drawn. In addition, given that massed study improves memory within a short retention interval (e.g., Balota, Duchek, & Paullin, 1989), extending the retention interval would also likely be effective.

It will also be important to test the generality of our results using additional item-specific (e.g., familiarity ratings) and relational (e.g., integrative imagery) tasks. This future direction is also important given our two item-specific tasks, and our two relational tasks did not produce the parallel outcomes when a single block was used (in Experiment 2, cf. Experiment 1). Task analyses could also be useful for predicting the relative amount of item-specific vs. relational processing that a given task will produce. Such analyses could also be useful for identifying how effective different item-specific and relational processing subtypes are at producing EV benefits.

Increasing the number of study blocks would also be informative. With one exception (Maskarinec & Thompson, 1976), published studies have yet to examine EV effects using more than two blocks. There may be an ideal ratio of item-specific to relational tasks across blocks for maximizing EV benefits. For example, completing several relational processing tasks and one item-specific task may result in a larger EV benefit than completing several item-specific tasks and one relational task, or vice versa, depending on list type. This remains an open question.

A critical question is whether EV benefits are driven by how participants encode information at study (e.g., increased elaborative encoding) and/or by how they utilize that information at test (e.g., increased memory monitoring). Traditionally, EV refers to variations at encoding, but it does not specify whether the benefits of EV are due to processes that occur at encoding and/or at retrieval. The separation of these loci in various memory paradigms has received considerable attention from researchers, and includes use of signal-detection indices (Gunter, Bodner, & Azad, 2007; Huff & Bodner, 2013), within-subjects designs (Schacter, Israel, & Racine, 1999), and inclusion tests (Hege & Dodson, 2004; Hunt et al., 2011).

Finally, irrespective of whether the locus of EV benefits is at encoding and/or at retrieval, EV may be of great utility for students and educators alike. This important potential warrants exploration. We found that EV benefits were restricted to weakly related word lists and variable-processing conditions, but future research should examine whether EV benefits extend to educationally relevant materials. Can item-specific or relational processing be implemented when reading paragraphs of text for meaning, learning mathematical formulas, or memorizing historical facts? Do students already use such techniques? Do the materials need to be weakly related (and if so, what *are* weakly related study materials, functionally)? Adapting the EV paradigm toward these questions provides a profitable avenue of research.

Conclusion

The present study aimed to reinvigorate research on EV as a means of improving memory. Three types of EV benefit were identified. Limited EV benefits occurred when the study task, but not the type of processing, varied across study experiences. More robust EV benefits occurred when both the tasks and the type of processing the tasks required varied across study experiences. A third type of EV benefit occurred when the type of processing sponsored by a task and the type of processing sponsored by the study list were complementary rather than redundant with each other. For researchers, our study provides new support for an old hypothesis that has long been applied to a variety of memory phenomena. For students and educators, we hope that our study inspires the development of new applications for enhancing memory.

Acknowledgments

This study comprises the first author's doctoral dissertation and was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) Postgraduate Scholarship and by NIH grant 5T32AG000030-38 to M.J.H. and by an NSERC Discovery Grant to G.E.B. M.J.H. is now at Washington University in St. Louis. We extend special thanks to Sara Davis, Aram Lee, and Erin Shumlich for assistance in data collection.

Appendix A

See Tables A1–A4.

Appendix B

B.1. Experiments 1 and 2: Clustering analyses

False recognition was lower following item-specific than relational tasks in both Experiments 1 and 2, consistent with our tasks sponsoring the expected type of processing. Here an additional method for ensuring that the two processing types were manipulated as intended was explored, namely, using adjusted-ratio-of-clustering (ARC) indices (Roenker, Thompson, & Brown, 1971; Senkova & Otani, 2012). ARC scores measure of the extent to which items reported during a free recall task are organized by category membership (0 = chance clustering, 1 = perfect clustering). ARC scores can be greater following relational than item-specific tasks (Einstein & Hunt, 1980), and have also been found to correlate with retention (Mulligan, 2005).

An ARC score was calculated for each participant in each experiment. An inclusion requirement was that a participant recalled at least one item from each of the four categories. One MI group participant in Experiment 2B did not meet this requirement and was excluded, leaving 447 mean ARC scores for analysis (see Table B1).

The MI/MI, PR/PR, MI/PR, MI, and PR groups across experiments formed the pooled IS group, and the CS/CS, NC/NC, CS/NC, CS, and NC groups formed the pooled REL group. A 2 (processing type: IS vs. REL) $\times$ 2 (list type: strongly related vs. weakly related) between-subjects ANOVA was then conducted on the mean ARC scores. This analysis revealed a significant main effect of processing type, $F(1,315) = 25.10$, $MSE = .04$, $\eta_p^2 = .07$. Contrary to expectations, ARC scores were *greater* in the IS than the REL group (.35 vs. .23). A main effect of list type was also found, $F(1,315) = 6.41$, $MSE = .04$, $\eta_p^2 = .02$, confirming that ARC scores were greater with strongly related lists than weakly related lists (.32 vs. .26). The interaction was not significant, $F(1,315) = 1.32$, $MSE = .04$, $p = .25$.

ARC scores were also expected to be greater for groups that produced (vs. did not produce) an EV benefit, based on the finding that clustering improves retention (Mulligan, 2005). Therefore, ARC scores were expected to be greater in the variable-processing groups relative to the variable-task and repeated-task groups with weakly related lists (Experiments 1B and 2B), but not with strongly related lists (Experiment 1A and 2A). A 3 (group: repeated task vs. variable task vs. variable processing) $\times$ 2 (list type: strongly related vs. weakly related) between-subjects ANOVA revealed only a main effect of list type, $F(1,314) = 5.89$, $MSE = .04$, $\eta_p^2 = .02$ (other $Fs < 1.62$, $ps > .20$). ARC scores were greater for strongly (vs. weakly) related lists (.34 vs. .29). Variable processing did not result in greater clustering.

As expected, clustering was greater following strongly (vs. weakly) related lists, but unexpectedly, ARC scores

Table A1
Experiments 1A and 1B: mean (SD) proportion of recalled list items and critical items, and average number of extra-list intrusions.

Group/tasks	Exp. 1A: Strongly related lists			Exp. 1B: Weakly related lists		
	List items	Critical items	Intrusions	List items	Critical items	Intrusions
Repeated task	.38 (.10)	.01 (.02)	.19 (.61)	.36 (.07)	.00 (.02)	.33 (.68)
IS–IS (MI/MI)	.42 (.08)	.01 (.03)	.06 (.25)	.33 (.07)	.00 (.00)	.13 (.34)
IS–IS (PR/PR)	.39 (.09)	.01 (.02)	.06 (.25)	.38 (.06)	.00 (.00)	.19 (.54)
REL–REL (CS/CS)	.33 (.09)	.02 (.03)	.13 (.34)	.37 (.06)	.01 (.02)	.25 (.44)
REL–REL (NC/NC)	.39 (.14)	.01 (.02)	.50 (1.10)	.35 (.09)	.01 (.03)	.75 (1.00)
Variable task	.40 (.10)	.01 (.02)	.16 (.51)	.34 (.08)	.00 (.02)	.53 (1.14)
IS–IS (MI/PR)	.44 (.10)	.00 (.01)	.00 (.00)	.36 (.08)	.00 (.00)	.31 (.49)
REL–REL (CS/NC)	.36 (.10)	.02 (.03)	.31 (.70)	.33 (.08)	.01 (.02)	.75 (1.49)
Variable processing	.40 (.08)	.02 (.04)	.27 (.98)	.39 (.08)	.01 (.02)	.19 (.56)
IS–REL (MI/CS)	.40 (.06)	.03 (.06)	.69 (1.91)	.39 (.09)	.00 (.02)	.31 (.79)
IS–REL (MI/NC)	.44 (.09)	.02 (.04)	.13 (.38)	.39 (.12)	.01 (.02)	.06 (.25)
IS–REL (PR/CS)	.40 (.09)	.02 (.03)	.00 (.00)	.38 (.07)	.01 (.02)	.31 (.70)
IS–REL (PR/NC)	.38 (.07)	.02 (.03)	.25 (.45)	.40 (.06)	.01 (.02)	.06 (.25)

Table A2
Experiments 1A and 1B: mean (SD) proportion of "old" recognition responses to correctly studied list items (hits), nonstudied control items (false alarms), nonstudied critical items (CI false alarms), nonstudied critical control items (false alarms).

Group/tasks	Exp. 1A: Strongly related lists				Exp. 1B: Weakly related lists			
	List items	List items controls	Critical items	Critical items controls	List items	List items controls	Critical items	Critical items controls
Repeated task	.91 (.08)	.01 (.02)	.19 (.17)	.01 (.02)	.93 (.07)	.02 (.03)	.11 (.10)	.02 (.04)
IS–IS (MI/MI)	.97 (.03)	.01 (.01)	.13 (.18)	.01 (.02)	.95 (.05)	.01 (.02)	.06 (.06)	.01 (.02)
IS–IS (PR/PR)	.92 (.03)	.02 (.02)	.17 (.14)	.01 (.03)	.95 (.06)	.02 (.03)	.11 (.10)	.02 (.03)
REL–REL (CS/CS)	.90 (.09)	.01 (.03)	.31 (.17)	.01 (.03)	.93 (.08)	.01 (.02)	.12 (.11)	.02 (.03)
REL– REL (NC/NC)	.85 (.07)	.01 (.02)	.17 (.14)	.01 (.02)	.87 (.07)	.03 (.03)	.13 (.09)	.04 (.05)
Variable task	.94 (.04)	.01 (.03)	.25 (.25)	.02 (.05)	.93 (.07)	.03 (.04)	.14 (.16)	.04 (.06)
IS–IS (MI/PR)	.96 (.04)	.01 (.03)	.14 (.16)	.00 (.01)	.96 (.06)	.02 (.03)	.09 (.08)	.03 (.06)
REL–REL (CS/NC)	.93 (.04)	.01 (.03)	.36 (.27)	.04 (.07)	.90 (.07)	.04 (.05)	.19 (.20)	.04 (.05)
Variable processing	.94 (.05)	.02 (.03)	.28 (.22)	.04 (.07)	.95 (.05)	.02 (.03)	.13 (.13)	.02 (.04)
IS–REL (MI/CS)	.94 (.05)	.02 (.04)	.22 (.18)	.04 (.07)	.97 (.03)	.02 (.02)	.11 (.11)	.01 (.03)
IS–REL (MI/NC)	.94 (.07)	.03 (.05)	.35 (.28)	.05 (.12)	.94 (.06)	.02 (.03)	.15 (.12)	.04 (.05)
IS–REL (PR/CS)	.95 (.05)	.01 (.02)	.28 (.19)	.03 (.04)	.95 (.04)	.02 (.02)	.16 (.17)	.01 (.02)
IS–REL (PR/NC)	.94 (.04)	.02 (.02)	.27 (.19)	.03 (.05)	.95 (.05)	.04 (.07)	.15 (.12)	.04 (.05)

were greater following item-specific (vs. relational) tasks (cf. Einstein & Hunt, 1980). This pattern is also divergent from demonstrations that item-specific processing reduces associative false recognition (e.g., Huff & Bodner, 2013; Hunt et al., 2011).

However, a closer look at other studies reveals that greater clustering for item-specific (vs. relational) processing is not unprecedented. Hodge and Otani (1996) reported that item-specific PR and MI tasks showed greater ARC scores than relational tasks such as NC and relational imagery. Similarly, Hunt and Seta (1984, Experiment 1) examined ARC clustering for item-specific and relational tasks when studying lists of items that differed in terms of the number of exemplars studied from a category set. Lists ranged from 2 to 16 exemplars per category set. When 2 exemplars from each category were study, ARC scores were greater following a relational (vs. item-specific) task. Conversely, when the number of exemplars from each category increased from 4 to 16, item-specific tasks resulted in equivalent or numerically greater clustering scores than relational tasks. The authors stated that larger category set sizes are more accessible in memory than smaller sets and once a category was accessed, item-specific information is required to recall individual items, inflating clustering measures. Our study utilized 20 exemplars from each category set which likely enhanced category accessibility.

Table A3
Experiments 2A and 2B: mean (SD) proportion of list items correctly recalled, critical items falsely recalled, and average number of extra-list intrusions.

Group/tasks	Exp. 1A: Strongly related lists			Exp. 1B: Weakly related lists		
	List items	Critical items	Intrusions	List items	Critical items	Intrusions
Item-specific processing						
Mental imagery	.37 (.12)	.00 (.01)	.19 (.40)	.25 (.08)	.01 (.02)	.69 (.79)
Pleasantness ratings	.29 (.09)	.04 (.05)	.44 (1.03)	.24 (.08)	.02 (.05)	.94 (1.39)
Relational processing						
Category sorting	.28 (.06)	.05 (.07)	.31 (.48)	.31 (.06)	.00 (.01)	.44 (.73)
Narrative construction	.27 (.09)	.02 (.03)	.75 (.93)	.22 (.12)	.00 (.00)	2.13 (2.85)

Table A4
Experiments 2A and 2B: mean (SD) proportion of "old" recognition responses to correctly studied list items (hits), nonstudied control items (false alarms), nonstudied critical items (CI false alarms), nonstudied critical control items (false alarms).

Group/task	Exp. 1A: Strongly related lists				Exp. 1B: Weakly related lists			
	List items	List items controls	Critical items	Critical items controls	List items	List items controls	Critical items	Critical items controls
Item-specific tasks								
Mental imagery	.93 (.05)	.02 (.03)	.15 (.10)	.03 (.06)	.89 (.08)	.05 (.06)	.12 (.09)	.05 (.09)
Pleasantness ratings	.87 (.14)	.04 (.09)	.24 (.16)	.06 (.07)	.85 (.14)	.03 (.04)	.12 (.10)	.03 (.05)
Relational tasks								
Category sorting	.83 (.10)	.02 (.04)	.28 (.18)	.02 (.08)	.88 (.08)	.02 (.02)	.24 (.22)	.02 (.05)
Narrative construction	.82 (.11)	.02 (.03)	.26 (.16)	.04 (.07)	.78 (.16)	.17 (.18)	.31 (.23)	.21 (.24)

Table B1
Adjusted-Ratio-of-Clustering (ARC) scores (SD) for Experiments 1 and 2.

Group/task(s)	Strongly related lists	Weakly related lists
Repeated tasks (Exp. 1)	.31 (.21)	.24 (.23)
IS–IS (MI–MI)	.46 (.16)	.32 (.18)
IS–IS (PR–PR)	.34 (.18)	.26 (.23)
REL–REL(CS–CS)	.30 (.24)	.37 (.13)
REL–REL (NC–NC)	.15 (.16)	.03 (.21)
Variable tasks (Exp. 1)	.38 (.23)	.28 (.21)
IS–IS (MI–PR)	.50 (.21)	.38 (.21)
REL–REL(CS–NC)	.27 (.19)	.19 (.19)
Variable processing (Exp. 1)	.34 (.21)	.32 (.18)
IS–REL (MI–NC)	.27 (.19)	.24 (.15)
IS–REL (MI–CS)	.45 (.19)	.44 (.18)
IS–REL (PR–NC)	.28 (.18)	.24 (.16)
IS–REL (PR–CS)	.37 (.23)	.35 (.17)
Single task (Exp. 2)	.29 (.23)	.24 (.23)
IS (MI)	.39 (.22)	.32 (.18)
IS (PR)	.27 (.24)	.26 (.23)
REL (CS)	.31 (.25)	.37 (.13)
REL (NC)	.21 (.22)	.03 (.22)

Thus it is perhaps unsurprising that item-specific tasks resulted in greater clustering.

These findings challenge the utility of using ARC clustering scores as a metric of item-specific and relational processing, at least when category set size is large. We suggest that other measures may provide better converging evidence of processing type. These include associative false memories (which we relied on in here) and cumulative recall curves (Burns & Hebert, 2005; Burns & Schoff, 1998) that assess item gains and losses over repeated tests (our use of a single 5 min recall test precluded a cumulative recall analysis). With these measures, item-specific and relational processing tasks such as ours produce consistent, expected patterns.

References

Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior, 22*, 261–295. http://dx.doi.org/10.1016/S0022-5371(83)90201-3.

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. http://dx.doi.org/10.1037/0882-7974.4.1.3.

Balota, D. A., Duchek, J. M., & Logan, J. M. (2007). Is expanded retrieval practice a superior form of spaced retrieval? A critical review of the extent literature. In Nairne, J. S. (Ed.), *The foundations of remembering: Essays in honor of Henry L. Roediger III* (pp. 83–105).

Balota, D. A., & Lorch, R. F. (1986). Depth of automatic spreading activation: Mediated priming effects in pronunciation but not lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*, 336–345. http://dx.doi.org/10.1037/0278-7393.12.3.336.

Battig, W. F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monograph, 80*, 1–45. http://dx.doi.org/10.1037/h0027577.

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

Bobrow, S. A. (1970). Memory for words in sentences. *Journal of Verbal Learning and Verbal Behavior, 9*, 363–372. http://dx.doi.org/10.1016/S0022-5371(70)80074-3.

Bower, G. H., Lesgold, A. M., & Tieman, D. (1969). Grouping operations in free recall. *Journal of Verbal Learning and Verbal Behavior, 8*, 481–493. http://dx.doi.org/10.1016/S0022-5371(69)80092-7.

Burns, D. J., Burns, S. A., & Hwang, A. J. (2011). Adaptive memory: Determining the proximate mechanisms responsible for the memorial advantage of survival processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 206–218. http://dx.doi.org/10.1037/a0021325.

Burns, D. J., & Hebert, T. (2005). Using cumulative-recall curves to assess the extent of relational and item-specific processing. *Memory, 13*, 189–199. http://dx.doi.org/10.1080/09608210344000670.

Burns, D. J., & Schoff, K. M. (1998). Slow and steady often ties the race: Effects of item- specific and relational processing cumulative recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 1041–1051. http://dx.doi.org/10.1037/0278-7393.24.4.1041.

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. http://dx.doi.org/10.1037/0033-2909.132.3.354.

Chwilla, D. J., & Kolk, H. H. J. (2002). Three-step priming in lexical decision. *Memory & Cognition, 30*, 217–225. http://dx.doi.org/10.3758/BF03195282.

Collins, A. M., & Loftus, E. F. (1975). A spreading activation theory of semantic processing. *Psychological Review, 82*, 407–428. http://dx.doi.org/10.1037/0033-295X.82.6.407.

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

Dempster, F. N. (1987). Effects of variable encoding and spaced presentations on vocabulary learning. *Journal of Educational Psychology, 79*, 162–170. http://dx.doi.org/10.1037/0022-0663.79.2.162.

Einstein, G. O., & Hunt, R. R. (1980). Levels of processing and organization: Additive effects of individual-item and relational processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 6*, 588–598. http://dx.doi.org/10.1037/0278-7393.6.5.588.

Elmes, D. G., & Bjork, R. A. (1975). The interaction of encoding and rehearsal processes in the recall of repeated and nonrepeated items. *Journal of Verbal Learning and Verbal Behavior, 14*, 30–42. http://dx.doi.org/10.1016/S0022-5371(75)80004-1.

Engelkamp, J., Biegelmann, U., & McDaniel, M. A. (1998). Relational and item-specific information: Trade-off and redundancy. *Memory, 6*, 307–333. http://dx.doi.org/10.1080/741942360.

Estes, W. K. (1950). Toward a statistical theory of learning. *Psychological Review, 57*, 94–107. http://dx.doi.org/10.1037/h0058559.

Galbriath, R. C. (1975). Encoding variability and the concreteness effect in paired-associate learning. *Memory & Cognition, 3*, 282–286.

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

Glenberg, A. M. (1977). Influences of retrieval processes on the spacing effect in free recall. *Journal of Experimental Psychology: Human Learning and Memory, 3*, 282–294. http://dx.doi.org/10.1037/0278-7393.3.3.282.

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

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

Gunter, R. W., Bodner, G. E., & Azad, T. (2007). Generation and mnemonic encoding induce a mirror effect in the DRM paradigm. *Memory & Cognition, 35*, 1083–1092. http://dx.doi.org/10.3758/BF031934080.

Hege, A. C. G., & Dodson, C. S. (2004). Why distinctive information reduces false memories: Evidence for both impoverished relational encoding and distinctiveness heuristic accounts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 787–795. http://dx.doi.org/10.1037/0278-7393.30.4.787.

Hintzman, D. L. (2004). Judgment of frequency versus recognition confidence: Repetition and recursive reminding. *Memory & Cognition, 32*, 336–350. 10.103758/BF03196863.

Hintzman, D. L. (1974). Theoretical implications of the spacing effect. In R. L. Solson (Ed.), *Theories in cognitive psychology: The Loyola symposium* (pp. 386). England: Lawrence Erlbaum.

Hintzman, D. L., & Stern, L. D. (1978). Contextual variability and memory for frequency. *Journal of Experimental Psychology: Human Learning and Memory, 4*, 539–549. http://dx.doi.org/10.1037/0278-73934.5.539.

Hodge, M. H., & Otani, H. (1996). Beyond category sorting and pleasantness rating: Inducing relational and item-specific processing. *Memory & Cognition, 24*, 110–115.

Huff, M. J., & Bodner, G. E. (2013). When does memory monitoring succeed versus fail? Comparing item-specific and relational encoding in the DRM paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. http://dx.doi.org/10.1037/90031338.

Huff, M. J., Coane, J. H., Hutchison, K. A., Grasser, E. B., & Blais, J. E. (2012). Interpolated task effects on directed and mediated false recognition: Effects of initial recall, recognition, and the ironic effect of guessing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*, 1720–1730. http://dx.doi.org/10.1037/a0028476.

Huff, M. J., Meade, M. L., & Hutchison, K. A. (2011). Age-related differences in guessing on free and forced recall tests. *Memory, 19*, 317–330. http://dx.doi.org/10.1080/09658211.2011.568494.

Hunt, R. R., & Einstein, G. O. (1981). Relational and item-specific information in memory. *Journal of Verbal Learning and Verbal Behavior, 20*, 497–514. http://dx.doi.org/10.1016/S0022-5371(81)90138-9.

Hunt, R. R., & McDaniel, M. A. (1993). The enigma of organization and distinctiveness. *Journal of Memory and Language, 32*, 421–445. http://dx.doi.org/10.1006/jmla.1993.1023.

Hunt, R. R., & Seta, C. E. (1984). Category size effects in recall: The roles of relational and individual item information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 454–464. http://dx.doi.org/10.1037/0278-7393.10.3.464.

Hunt, R. R., Smith, R. E., & Dunlap, K. R. (2011). How does distinctive processing reduce false recall? *Journal of Memory and Language, 65*, 378–389. http://dx.doi.org/10.1016/j.jml.2011.06.003.

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

Johnston, W. A., Coots, J. H., & Flickinger, R. G. (1972). Controlled semantic encoding and the effect of repetition lag on free recall. *Journal of Verbal Learning and Verbal Behavior, 11*, 784–788. http://dx.doi.org/10.1016/S0022-5371(72)80013-6.

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

Mandler, G. (1980). Rcognizing: The judgment of previous occurrence. *Psychological Review, 87*, 252–271. http://dx.doi.org/10.1037/0033-295X.87.3.252.

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

Maskarinec, A. S., & Thompson, C. P. (1976). The within-list distributed practice effect: Tests of the varied context and varied encoding hypotheses. *Memory & Cognition, 6*, 741–746. http://dx.doi.org/10.2307/1421519.

McCabe, D. P., Presmanes, A. G., Robertson, C. L., & Smith, A. D. (2004). Item-specific processing reduces false memories. *Psychonomic Bulletin & Review, 11*, 1074–1079.

McDaniel, M. A., & Einstein, G. O. (1989). Material-appropriate processing: A contextualist approach to reading and studying. *Educational Psychology Review, 1*, 113–145. http://dx.doi.org/10.1007/BF01326639.

McDaniel, M. A., Einstein, G. O., & Lollis, T. (1988). Qualitative and quantitative considerations in encoding difficulty effects. *Memory & Cognition, 16*, 8–14. http://dx.doi.org/10.3758/BF03197740.

McDaniel, M. A., & Masson, E. J. (1985). Altering memory representations through retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 371–385. http://dx.doi.org/10.1037/0278-7393.11.2.371.

McDaniel, M. A., Waddill, P. J., & Einstein, G. O. (1988). A contextual account of the generation effect: A three-factor theory. *Journal of Memory and Language, 27*, 521–536. http://dx.doi.org/10.1016/0749-596X(88)90023-X.

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. http://dx.doi.org/10.1016/S0022-5371(70)80107-4.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing vs. transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16*, 249–255. http://dx.doi.org/10.1016/S0022-5371(77)80016-9.

Mulligan, N. W. (2005). Total retrieval time and hypermnesia: Investigating the benefits of multiple recall tests. *Psychological Research, 69*, 272–284. http://dx.doi.org/10.1007/s00426-004-0178-5.

Postman, L., & Knecht, K. (1983). Encoding variability and retention. *Journal of Verbal Learning and Verbal Behavior, 22*, 133–152. http://dx.doi.org/10.1016/S0022-5371(83)90101-9.

Roediger, H. L., III, Sanches, J. B., & Agarwal, P. K. (2011). Does variable encoding affect learning and retention relative to constant encoding? *Paper presented at the 52nd annual meeting of the Psychonomic Society*, Seattle, WA.

Roediger, H. L., III, & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*, 249–255. http://dx.doi.org/10.1111/j.1467-9280.2006.01693.x.

Roediger, H. L., III, & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 803–814. http://dx.doi.org/10.1037/0878-7393.21.4.803.

Roenker, D. L., Thompson, C. P., & Brown, S. C. (1971). Comparison of measures for the estimation of clustering in free recall. *Psychological Bulletin, 76*, 45–48. http://dx.doi.org/10.1037/h0031355.

Schacter, D. L., Israel, L., & Racine, C. (1999). Suppressing false recognition in younger and older adults: The distinctiveness heuristic. *Journal of Memory and Language, 40*, 1–24. http://dx.doi.org/10.1006.jmla.1998.2611.

Senkova, O., & Otani, H. (2012). Category clustering calculator for free recall. *Advances in Cognitive Psychology, 8*, 292–295. http://dx.doi.org/10.2478/v10053-008-0124-y.

Toglia, M. P., Neushatz, J. S., & Goodwin, K. A. (1999). Recall accuracy and illusory memories: When more is less. *Memory, 7*, 233–256. http://dx.doi.org/10.1080/741944069.

Tulving, E. (1962). Subjective organization in free recall of "unrelated" words. *Psychological Review, 69*, 344–354. http://dx.doi.org/10.1037/h0043150.

Tulving, E. (1966). Subjective organization and effects of repetition in multi-trial free-recall learning. *Journal of Verbal Learning and Verbal Behavior, 5*, 193–197. http://dx.doi.org/10.1016/S0022-5371(66)80016-6.

Tulving, E. (1968). When is recall higher than recognition? *Psychonomic Science, 10*, 53–54.

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

Van Overschelde, J. P., Rawson, K. A., & Dunlosky, J. (2004). Category norms: An updated and expanded version of the Battig and Montague (1969) norms. *Journal of Learning and Memory, 50*, 289–335. http://dx.doi.org/10.1016/j.jml.2003.10.003.

Williams, R. F., & Underwood, B. J. (1970). Encoding variability: Tests of the Martin hypothesis. *Journal of Experimental Psychology, 86*, 317–324. http://dx.doi.org/10.37/h0029998.

Winograd, E., & Geis, M. F. (1974). Semantic encoding and recognition memory: A test of the encoding variability theory. *Journal of Experimental Psychology: General, 102*, 1061–1068. http://dx.doi.org/10.1037/h0036386.

Young, D. R., & Bellezza, F. S. (1982). Encoding variability, memory organization, and the repetition effect. *Journal of Experimental Psychology: Learning, Memory and Cognition, 8*, 545–559. http://dx.doi.org/10.1037/0278-7393.8.6.545.