On the memory benefits of repeated study with variable tasks

Zawadzka, Katarzyna; Baloro, Samantha; Wells, Jennifer; Wilding, Edward L.; Hanczakowski, Maciej

DOI:
10.1037/xlm0001013

Citation for published version (Harvard):
Zawadzka, K, Baloro, S, Wells, J, Wilding, EL & Hanczakowski, M 2021, 'On the memory benefits of repeated study with variable tasks', *Journal of Experimental Psychology: Learning Memory and Cognition*, vol. 47, no. 7, pp. 1067-1082. https://doi.org/10.1037/xlm0001013

On the memory benefits of repeated study with variable tasks

Katarzyna Zawadzka[1], Samantha Baloro[2], Jennifer Wells[2], Edward L. Wilding[3], and Maciej Hanczakowski[1]

[1] SWPS University, Poland

[2] University of Nottingham, UK

[3] University of Birmingham, UK

Author Note

Katarzyna Zawadzka, Interdisciplinary Center for Applied Cognitive Studies, SWPS University; Samantha Baloro and Jennifer Wells, School of Psychology, University of Nottingham; Edward L. Wilding, School of Psychology, University of Birmingham; Maciej Hanczakowski, Interdisciplinary Center for Applied Cognitive Studies, SWPS University.

All materials and data are available at https://osf.io/a72q4/

The authors would like to thank Ewa Butowska, Karolina Lukasik, Kalina Kordyńska, Krzysztof Piątkowski, and Oliwia Zaborowska for their assistance with data collection for Experiments 5-7.

Correspondence should be addressed to Katarzyna Zawadzka or Maciej Hanczakowski, Interdisciplinary Center for Applied Cognitive Studies, SWPS University, ul. Chodakowska 19/31, 03-815, Warszawa, Poland, email: k.n.zawadzka@gmail.com or maciej.hanczakowski@gmail.com

Abstract

Encoding variability refers to the situation in which repeated items are processed in different ways on each presentation. Superior memory performance resulting from encoding variability is sometimes argued to underlie important phenomena in human memory such as the spacing effect. However, the memory benefits of encoding variability are often elusive. Here we investigated encoding variability in ten experiments in which participants studied words with the same or different orienting tasks across repetitions. We have found the benefits of variable encoding to depend upon the number of learning cycles and the retrieval demands at test. These results are interpreted in light of a distinction between different components of memory representations established at study, suggesting that encoding variability promoted via different orienting tasks -as implemented in the present study – fosters more elaborate encoding of semantic features. This augmented semantic component benefits memory performance only when a memory test is utilized that taps predominantly semantic features of memory representations, minimizing the role of contextual and relational factors.

Keywords: recall; encoding variability; component-levels theory

On the memory benefits of repeated study with variable tasks

When confronted with difficult materials to learn, people tend to restudy these materials in repeated cycles of learning. A question of practical importance is how to make such repetitions as effective as possible. One seemingly obvious recommendation is to elaborate study materials in order to create rich memory traces. Elaborative encoding supports later remembering, either simply because richly encoded items are more likely to be retrieved (Bradshaw & Anderson, 1982; Karpicke & Smith, 2012), or because they support a variety of monitoring strategies that ensure minimal interference from other contents of memory (Gallo, Meadow, Johnson, & Foster, 2008; Hanczakowski & Mazzoni, 2011; Huff, Bodner, & Fawcett, 2015). However, theories of learning suggest also the importance of *varied* learning. According to the encoding variability hypothesis, memory performance is maximized if, during repeated study of the same materials, each occurrence of the item is accompanied by either different cognitive operations or different contextual features (see Estes, 1955, for an early formulation). Variable encoding is thought to go beyond standard elaboration by establishing even richer memory traces – with each encoding opportunity adding different features to the stored memory trace – which in turn provides many routes by which a given item can be accessed subsequently (Glenberg, 1979).

Although the principle of variable encoding is relevant to educational practice, research devoted to educational applications of basic memory principles does not foreground it. For example, a recent overview of effective learning strategies by Weinstein, Madan, and Sumeracki (2018) covers several well-established methods of how to increase memory retention in applied settings – spacing of practice sessions, interleaving, retrieval practice, elaboration, concrete examples, dual coding – but encoding variability is not explicitly discussed and remains only tentatively related to some of the presented methods. Thus, for example, although dual coding may involve repeated learning in variable modalities, it can also be instantiated in a single learning session. Similarly, additional elaboration may be involved when learning proceeds in variable modes – when varied encoding is understood as further elaboration of an existing memory representation – but it can also be achieved by deeper repeated encoding in a single study session or by repeated elaboration under constant encoding conditions. Furthermore, encoding variability may actually underlie the benefits of spacing – as discussed shortly – but there are also other accounts of the spacing effect that do not postulate the involvement of encoding variability.

This lack of interest in encoding variability in applied studies is mirrored in a relative scarcity of basic empirical work concerning the effects of encoding variability. Although the principle of encoding variability remains embedded in important theoretical models of memory, of which the temporal context model (Howard & Kahana, 2002) is perhaps the most prominent, there are also recent theoretical proposals (Benjamin & Tullis, 2010) and empirical studies (Xue, Dong, Chen, Lu, Mumford, & Poldrack, 2010) according to which the concept of encoding variability should be abandoned as an explanatory tool when repeated study is considered.

The strengths and weaknesses of the encoding variability hypothesis have often been considered in relation to the phenomenon of spacing, by which to-be-learned materials are better remembered if their consecutive presentations are separated in time, as compared to massed presentations (see Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006; Delaney, Verkoeijen, & Spirgel, 2010 for reviews). The encoding variability account of the spacing effect proposes that temporal context – the entirety of all features peripheral to to-be-learned materials, such as one's thoughts, mood, and features of the physical environment at the time of encoding – changes continuously during study and the features of temporal context accompanying presentations of studied items are stored in memory traces. With spaced presentations, each study opportunity allows for encoding different features due to contextual drift. With massed presentations, by contrast, a limited number of contextual features are stored. Thus, spaced encoding is characterized by greater variability of stored contextual features, which – consistent with the encoding variability hypothesis – results in more effective subsequent memory retrieval.

The problem with assessing the encoding variability hypothesis by varying spacing across repetitions is one of experimental control – there is very little control the experimenter can exercise over variability of features that depends on unobservable changes in temporal context. A more direct test of the encoding variability hypothesis requires either a manipulation of orienting tasks or the way the stimuli are presented that would ensure that repeated presentations of items are either encoded in the same way or variably, preferably with spacing equated across the conditions being compared. A number of such experiments directly manipulating the type of encoding have been conducted, with experimental methods used to induce variable encoding including presenting homographs with their various meanings primed by preceding items (Slamecka & Barlow, 1979), inclusion of to-be-remembered words in different sentences (D'Agostino & DeRemer, 1973; Postman & Knecht, 1983), presentations of words in one or two languages known to bilingual participants (Glanzer & Duarte, 1971), or presenting to-be-remembered items in different environmental contexts (Smith, Glenberg, & Bjork, 1978). However, many of these studies failed to find the predicted benefits of varied over constant encoding or found them under limited conditions such as those involving explicit retrieval rather than repeated study (Smith & Handy, 2014). Of most interest to the present study, however, are experiments which manipulated encoding variability by using either repeated or changed orienting tasks across study presentations. Young and Bellezza (1982, Experiment 4; see also Hintzman & Stern, 1978; Bird, Nicholson, & Ringer, 1978; Rose, 1984) presented participants with a list of words for which they were asked to perform one of two orienting tasks: judging the size of an item or rating it for pleasantness. All words were presented twice and the orienting task for each of the words was either kept constant on both learning trials or switched. This enabled control over cognitive operations performed when study items were repeatedly presented, by either keeping them constant or varying them. In so doing, it provided proper grounds for assessing the effectiveness of encoding variability in supporting memory performance. However, when Young and Bellezza used this particular method they observed better memory for words studied twice with the same orienting task, singularly failing to support the encoding variability hypothesis.

The diverse pattern of empirical findings concerning the issue of encoding variability is presumably responsible for the disappearance of this hypothesis from the active research agenda in recent times. However, given the importance of this concept for theories of memory, we argue that these discrepant findings should serve as a spur for further empirical work and theory development. Why is it that the benefits of encoding variability are so elusive? One potential answer to this question has been offered by Huff and Bodner (2014), who proposed a distinction between variability in terms of tasks or processes. They suggested a link between the concept of encoding variability and the literature on different types of encoding processes that stress either the elaboration of the items themselves – termed *item-specific processing* – or the elaborations of relations between items of which the entire study materials consist – termed *relational processing* (Einstein & Hunt, 1980; Hunt & Einstein, 1981). According to the variable-processing version of the encoding variability hypothesis proposed by Huff and Bodner, the benefits of encoding variability are most likely to emerge when processing changes across item repetitions. That is, these benefits emerge when item-specific processing is engaged during one study session and relational processing is engaged during another study session. Variable tasks on their own are, according to Huff and Bodner, insufficient to promote better memory. That is, in their view two different tasks that require item-specific (or, alternatively, relational) processing in both study sessions should not yield better memory performance compared to a single task employed at both study sessions.

Some of the ideas proposed by Huff and Bodner (2014) were presaged by a theoretical framework applying the concept of encoding variability to explaining the spacing effect that was proposed by Glenberg (1979). In his component-levels theory, Glenberg proposed that at encoding three components of a memory representation are established (see Raaijmakers, 2003, for a computational model implementing this distinction). In Glenberg's words (p. 96): "Contextual components are automatically encoded upon the presentation of a to-be-remembered (TBR) item, and they represent the context in which the item is presented. Structural components encode the relationship discovered or imposed among sets of TBR items as a result of active processing. Descriptive components encode the specific item. The type of descriptive components encoded depends on the cognitive processes being used in the task." In this formulation, structural components are similar to the features of a memory representation established as a result of relational processing, while descriptive components are similar to the features established as a result of item-specific processing. The typology proposed by Glenberg adds also a contextual component that plays such a prominent role in the encoding variability accounts of the spacing effect. Reformulating the ideas of Huff and Bonder in the language of Glenberg's theory, encoding variability should benefit memory to the extent to which various components – structural, descriptive, contextual – are elaborated at different study opportunities. By contrast, elaboration of the same component across presentations, even if resulting from various tasks, should not benefit memory. Although the component-levels theory provides a language in which the variable-processing ideas of Huff and Bodner (2014) can be conveniently described, it is vital to note that this theory differs in its predictions concerning conditions under which encoding variability is likely to benefit memory. While Huff and Bodner argue for the necessity of encoding *different components*, Glenberg's theory proposes that encoding variability can benefit memory performance when *the same component* is encoded across repetitions – the conditions of variable tasks rather than variable processing.

The central tenet of the component-levels theory of Glenberg (1979) is its focus on the role of encoding-retrieval match. Briefly, Glenberg proposed that not only do encoding tasks differ in terms of the component which they promote, but – crucially – so do retrieval tasks in terms of the component they require. Thus, for example, free recall depends mostly on contextual and structural components, with contextual features serving as the first retrieval cue and already retrieved items cueing additional memory representations (see Raaijmakers & Shiffrin, 1981). Due to its great reliance on contextual features, free recall in the component-levels theory is sensitive to the richness of contextual features embedded in memory representations. Consequently, variable encoding that results in establishing memory representations that are rich in contextual features – such as encoding under spaced conditions – leads to memory benefits in a free recall task. While we return to the issue of encoding-retrieval match later in this article, it is vital to stress now that the component-levels theory clearly predicts the benefits of encoding variability as long as an appropriate test is used, even when processing is held constant across study presentations.

The purpose of the present study was to revisit the topic of encoding variability and its putative role in supporting memory retention. Faced with discrepant findings that have accumulated over many decades of research on this specific theoretical concept, we designed our study so that it could provide a strong basis for concluding whether benefits of variable encoding can be reliably observed. The theoretical framework used to guide the design and the interpretation of our experiments is the component-levels theory of Glenberg (1979). Specifically, we were interested in whether variable encoding confers benefits for memory retention also when features encoded across study presentations were restricted to a single component of memory representations. We focused on the descriptive component – the features constituting the representation of a to-be-remembered item itself, rather than its associations with context or other studied items – which we targeted by using orienting questions highlighting various aspects of the studied words. Our central manipulation involved either asking the same orienting question for a given item (constant encoding) or changing orienting questions across study presentations (varied encoding). Consistent with the encoding variability hypothesis, we generally predicted better memory retention for varied compared to constant encoding.

The following experiments can be arranged into two sets. The aim of Experiments 1a-3b was to establish conditions under which the benefits of encoding variability would reliably emerge. To anticipate, these conditions require a particular number of study word presentations followed – as predicted on the grounds of component-levels theory (Glenberg, 1979) – by an appropriate test. Experiments 4-7 aimed both to demonstrate the reliability those benefits and establish a set of boundary conditions for their occurrence.

Experiments 1a and 1b

In Experiments 1a and 1b, we followed the design of Young and Bellezza (1982, Experiment 4), examining the effects of encoding variability using a limited pool of orienting questions during study – either kept constant or switched across study presentations – and a free recall test to assess memory performance. Two versions of the experiment were conducted. Each was conducted by a different experimenter, at a different university, with small variations in the procedural details. The experiments differed in the number of repetitions of study words. In Experiment 1a, two cycles of learning, and thus two orienting questions were used. In Experiment 1b, three cycles of learning were used, necessitating the use of three orienting questions. It is worth noting that most of the investigations of encoding variability conducted to date limited the number of presentations of study materials to two, as examined here in Experiment 1a (but see Maskarinec & Thompson, 1976; Postman & Knecht, 1983, for exceptions). It is possible, however, that for memory benefits of encoding variability to emerge, a stronger manipulation is necessary. Using three different questions to encode words in the variable encoding condition in Experiment 1b tests this possibility.

Method

Participants. Forty students of the University of Nottingham (age range: 21-30, 8 males) participated in Experiment 1a, and 46 students of Cardiff University (age range: 18-23, 7 males) in Experiment 1b. All participants were fluent English speakers and received either payment or course credit for their participation.[1]

Materials and Design. Sixty-nine study words were chosen from the category norms developed by van Overschelde, Rawson, and Dunlosky (2004). All were concrete nouns containing four to eight letters and two to three syllables. Each word was taken from a different semantic category. From that word pool, 60 words were assigned to Experiment 1a, and further split into three lists of 20 words. Thirty-six words were assigned to Experiment 1b, and further split into two lists of 18 words. Each list was then assigned to a study-test block. In Experiment 1a, there were three study-test blocks, each of which consisted of two consecutive study phases separated by study instructions, followed by a free-recall test. There were two study-test blocks in Experiment 1b, each consisting of three consecutive study phases followed by a free-recall test.

A schematic overview of the study design is presented in Figure 1. A set of six orienting questions was created. Each single-sentence question queried a different aspect of a noun (e.g., 'Does it require electricity?', 'Could you paint it green?'). Two questions were assigned to each of the three study-test blocks of Experiment 1a and three questions were assigned to each of the two study-test blocks of Experiment 1b. At the beginning of each block the study list was randomly split into two halves, with half of each list being assigned to the varied-encoding condition and half to the constant-encoding condition. Each half was further divided by the number of encoding questions, and on the first study presentation each question was used for the same number of words in each

---

[1] Sample sizes differ across all experiments reported here for three reasons. First, Experiments 1a, 2a, and 3a had more trials per person than the comparison Experiments 1b, 2b, and 3b, as participants studied three unique word lists rather than two. Second, sample sizes were not determined before data collection as required for null-hypothesis testing. Instead, in the present investigation we planned to use Bayes factors to simultaneously assess evidence for the alternative hypothesis of differences between constant- and variable-encoding conditions and the null hypothesis, and in Bayesian analyses the posterior distribution does not depend on the stopping intentions (Kruschke & Liddell, 2018). Finally, the total number of participants in each experiment was also determined by the availability of participants (Simmons, Nelson, & Simonsohn, 2011): when the evidence seemed to reasonably favor one hypothesis over the other no new participants were recruited, but all the remaining participants signed up for the study were still tested.

condition (e.g., within each condition of Experiment 1a five words were presented with the first of the two orienting questions, and five with the second). In the constant-encoding condition, the same question was again used in all study phases. In the varied-encoding condition, the assignment of words to questions was changed for each study phase so that participants would see the words with each of the encoding questions assigned to that list: two in Experiment 1a, and three in Experiment 1b.

Both experiments were conducted in a within-participants design, with the only independent variable being the encoding condition. The encoding condition was manipulated within study lists. The assignment of words to the constant- and varied-encoding conditions was randomized anew for each participant. Constant encoding required the same word to be always presented with the same orienting question, while in the varied-encoding condition the orienting questions were different on each presentation of the word.

Procedure. Participants were tested either individually or in groups up to four, all of them at individual computers. At the start of the first study-test block, they were instructed to learn lists of words for a future test and, at the same time, provide yes/no answers to questions accompanying each word (by clicking on a ‘yes’ or ‘no’ on-screen button in Experiment 1a, or pressing Y or N in Experiment 1b). Each word was presented until the response to a question was given. After the first study phase, participants were informed that they would see the list again and that for some words the questions would be the same as in the first study phase, and for some the questions would be different. In Experiment 1b, the same instructions were also provided for the third study phase. The order in which the words were presented for study was randomized anew for each participant and for each study phase. This means that the mean spacing between the subsequent presentations of the same word was equal to the length of the study list. In Experiment 1a a study phase followed immediately a previous study phase, while in Experiment 1b there was a 1-min delay filled with simple math after each study phase. After the last study phase, a free-recall test followed. Participants were presented with a blank text box and asked to recall as many studied words as they could in 120 seconds. After that time, the procedure automatically terminated the test. Two or one further study-test blocks followed in Experiments 1a and 1b, respectively. The blocks were identical to the first one, bar the replacement of all study materials (words and orienting questions).

Results and Discussion

Descriptive statistics are presented in Table 1. For all of the analyses, we present the results of standard as well as Bayesian *t*-tests. Throughout the paper these two methods produced converging results, but we present both of them because they serve different purposes, with Bayesian statistics allowing us to assess evidence for both the alternative and the null hypotheses, and frequentist statistics facilitating comparisons to other studies. Given that the previous literature documented instances in which either varied- or constant-encoding condition was more beneficial for memory performance, in the present study we performed two-sided tests so our alternative hypothesis stated only that recall performance is not equal across encoding conditions. In both experiments, the difference in free-recall performance between the constant- and varied-encoding conditions was not significant, $t(39) = 1.132$, $p = .26$, $d = 0.18$ (Experiment 1a) and $t < 1$, $d = 0.09$ (Experiment 1b). Bayes factors were computed using the JASP software (JASP Team, 2018) with the 'default' prior of 0.707 centered at zero. Although Bayes factors are continuous measures of support for either the null or the alternative hypothesis, it is customary to treat Bayes factors exceeding the value of 3 as moderately supportive of either of the tested hypotheses as compared to the other (e.g., Lee & Wagenmakers, 2013). From this perspective, in neither Experiment 1a nor Experiment 1b was there any difference between the varied- and constant-encoding conditions in terms of free-recall performance, with the null hypothesis being 3.26 times more likely as the alternative hypothesis in Experiment 1a, and 5.30 times more likely in Experiment 1b.

Our findings agree with a number of earlier studies showing no benefits of changing encoding tasks across repeated cycles of learning (Bird et al., 1978; Postman & Knecht, 1983), and they remain consistent with the recent theoretical framework developed by Huff and Bodner (2014) which suggests that the benefits of variable encoding may be difficult to document when only variable-tasks conditions are employed.

Although these results do not support the idea of encoding variability benefits, it is premature to dismiss the possibility that variable encoding, even in the absence of variable processing, can result in superior memory performance. A potential problem with Experiments 1a and 1b, relevant also to previous studies using orienting tasks to manipulate constant versus variable encoding, is that variable encoding fostered by using only a limited sample of tasks does not seem particularly variable. In Experiment 1a, only two different tasks were used, in line with previous investigations of the same issue (e.g., Young & Bellezza, 1982). If all words are presented with one of only two orienting tasks, the features encoded for each item are bound to be very similar to features encoded for many other items, whether they serve in a constant- or variable-encoding condition. The issue was not much different in Experiment 1b, where only three different tasks were used.

The differentiation of encoded features across various studies items plays a fundamental role in some theories of memory. In the memory-as-discrimination framework (Nairne, 2002; see also

Beaman, Hanczakowski, Hodgetts, Marsh, & Jones, 2013; Goh & Lu, 2012), what supports memory performance is the extent to which a given to-be-remembered item can be tapped at the exclusion of all other encoded items. Applying this principle to the results of Experiments 1a and 1b, the use of a limited number of orienting tasks to foster encoding variability might have been ineffective in imbuing memory traces with unique features supporting subsequent discrimination. Indeed, the same argument can be derived from the component-levels theory which argues that the three components of memory representations – contextual, structural, descriptive – differ in terms of their generality, or the number of memory representations in which the same features are included: contextual features are typically the most general of the three (as all items are studied in the same experimental context), and descriptive features the most unique. This typical hierarchy of component levels is arguably disturbed when a limited number of orienting questions is used, resulting in an over-generality of the descriptive component and cue overload at retrieval (see Watkins & Watkins, 1975), thereby occluding the potential benefits of encoding variability. This issue was addressed in Experiments 2a and 2b, and then revisited in Experiment 7.

Experiments 2a and 2b

In this pair of experiments, we looked to foster unique processing on each study opportunity for items presented in the variable-encoding condition. This was done by using a unique orienting task for each of the studied items; in Experiments 2a and 2b the number of tasks was equal to the number of words in the study list. Additionally, items from the variable-encoding condition were encoded with a novel task on each repeated presentation, while items from the constant-encoding condition were – as before – encoded repeatedly with the same task. In this way, all studied items should become associated with unique features supporting subsequent retrieval, but items studied in the variable-encoding conditions should have more unique descriptive features than items from the constant-encoding conditions. According to the general version of the encoding variability hypothesis, memory performance should thus be augmented in the variable-encoding condition if variable tasks are in fact sufficient to produce the benefits of encoding variability.

Two versions of the experiment employing unique questions were again conducted. Both of them were revised versions of Experiments 1a and 1b, respectively, conducted with similar procedural details, by the same experimenters, and with new participants recruited from the same participant pools. Again, the major difference between experiments was the number of cycles of learning employed, with Experiment 2a using two presentations of each study item and Experiment 2b using three presentations.

Method

Participants. Twenty-six students of the University of Nottingham (age range: 18-33, 6 males) participated in Experiment 2a, and 40 students of Cardiff University (age range: 18-31, 5 males males) in Experiment 2b.

Materials, Design, and Procedure. The words used as study materials were the same as words used in Experiments 1a and 1b, with the exception that slightly longer lists were used in Experiment 2b, with 22 words per list. Additional words for this experiment were taken from the same pool as words for Experiment 1b. A list of 90 specific questions was created, including the six questions that were used in Experiments 1a and 1b. From this pool, in Experiment 2a there were 30 questions assigned to each of the three study-test blocks (with 20 used for the initial presentation and 10 used for the repetitions in the variable-encoding condition), and in Experiment 2b there were 44 questions assigned to each of the two study-test blocks (with 22 used for the initial presentation and 22 used for the two additional repetitions in the variable-encoding condition). The design of the experiment was the same as in Experiments 1a and 1b, with a single independent variable of constant versus variable encoding, manipulated within participants, and a dependent measure of correct free recall. The procedures for Experiments 2a and 2b were the same as in Experiments 1a and 1b, respectively.

Results and Discussion

Descriptive statistics are presented in Table 1. Again, the effect of variable encoding was not statistically significant, $t < 1$, $d = 0.01$ in Experiment 2a, and $t(40) = 1.023$, $p = .31$, $d = 0.16$ in Experiment 2b. The analysis of Bayes factors using JASP, with the same settings as in Experiments 1a and 1b, demonstrated that the use of unique encoding questions for each presentation of a study item did not change the pattern of results as compared to Experiments 1a and 1b, with the null hypothesis being 4.82 times more likely than the alternative hypothesis in Experiment 2a, and 3.64 times more likely in Experiment 2b.[2]

[2] A look at Table 2 suggests that performance in the present experiments was reduced compared to Experiments 1a and 1b. In the case of Experiments 1b and 2b, one contributing factor could be that the lists of words participants were asked to learn were longer in Experiment 2b (22 words) than Experiment 1b (18 words). However, for Experiments 1a and 2a the length of the study list was the same, yet a drop in performance seems evident. We suggest that this drop is likely to reflect differences in the ease of reinstating orienting questions as cues utilized at test. With few questions per list used in Experiments 1a (two questions) and 2a (three questions), such reinstatement should be relatively easy and thus participants could augment their performance with the use of these self-generated cues. However, with

We again failed to find support for the encoding variability hypothesis, this time under conditions that ensured that distinct features were encoded for all to-be-remembered items across study presentations. Together, the results presented so far seem to indicate that repeated encoding with variable tasks does not augment memory over and beyond encoding with the same task. However, there is still a caveat to this conclusion. So far, we have focused on examining encoding conditions only: all four experiments used solely a free-recall test to assess memory. However, as mentioned earlier, one of the main issues raised by the component-levels theory (Glenberg, 1979) is that revealing the benefits of encoding variability requires a test that taps the particular component of a memory representation that was enriched by variable processing. In the present study, we used an orienting task directed towards the descriptive component of memory representations and thus we should expect benefits of encoding variability to the extent to which memory tests also tap this specific descriptive component. It could be argued that free recall, with its reliance on contextual and structural components, is not ideally suited for revealing benefits of encoding variability at the descriptive level. For this reason, in Experiments 3a and 3b we turned to the question of the role of retrieval conditions.

Experiments 3a and 3b

In these experiments we sought to assess whether the benefits of variable encoding emerge in memory tests that tap predominantly the descriptive component, while minimizing the role of both structural and contextual components. Our manipulation of encoding variability was the same as in the previous experiments, with orienting questions directing participants' attention towards the semantic features of the concepts denoted by studied words. It is worth noting here that the descriptive component encompasses more than semantics and includes also other types of information such as phonemic or orthographic content; we revisit this issue in Experiment 6. Given, however, that here we varied semantic encoding during study, it was important to establish what would happen to memory performance if a test *specifically tapping semantic information* was used. To this aim, we used here the same design as in Experiments 2a and 2b, but we substituted free-recall tests with category-cued recall tests.

In terms of structural components, because in the present study we used study lists that contained a single exemplar for each of the semantic categories used, category cues should match

---

multiple questions used for Experiments 2a (30 questions) and 2b (44 questions), reinstatement of cues was likely to be less effective, reducing the benefits of such self-cueing.

only this single exemplar but not any of the other words, thus minimizing the role of any residual links across study words that participants could have created. Also, the order of category cues in the test was always randomized, further minimizing the role of inter-item associations possibly created on the basis of continuity at study presentations. In terms of contextual components, category cues were not included in the study phase and they thus constitute *extra-list cues* not seen at encoding and only introduced during the memory test. This minimizes the chance that categorical information becomes a conspicuous aspect of the encoding context. To the extent that encoding variability in terms of tasks serves to augment descriptive components, we would expect category-cued recall tests to be capable of revealing the role of encoding variability in supporting memory retention.

Again, two versions of the experiment employing category-cued recall tests were conducted. Both of them were revised versions of Experiments 2a and 2b, respectively, conducted with similar procedural details, by the same experimenters, and with new participants recruited from the same participant pools. The major difference between the experiments was again the number of cycles of learning employed, with Experiment 3a using two presentations of each study item and Experiment 3b using three presentations.

Method

Participants. Thirty-four students from the University of Nottingham (age range: 18-33, 15 males) took part in Experiment 3a, and 45 students from Cardiff University (age range: 18-34, 6 males) in Experiment 3b.

Materials, Design, and Procedure. The same study words and orienting questions were used as in Experiments 2a and 2b. A list of category names (e.g., “type of ship”, “mythical being”) was compiled from the category norms from which study words were originally taken (van Overschelde, Rawson, & Dunlosky, 2004). Thus, only one word on the study list belonged to each category. The design of the experiments was the same as in previous experiments, with a single independent variable of constant versus variable encoding, manipulated within participants, and a dependent measure of correct recall. The experimental procedure differed from that of Experiments 2a and 2b only in the type of the test completed at the end of each study-test block. A category-cued recall test was administered, with category names, which served as cues, presented in random order in separate rows on the screen. Participants were instructed that each category described one and only one of the studied words, and their task was to type in that word next to the category cue. As in previous experiments, the time for completing the test was 120 seconds, after which the procedure progressed automatically to the next screen.

Results and Discussion

Descriptive statistics are presented in Table 1. The pattern of results for Experiment 3a mirrored that from Experiments 1a and 2a. Despite the change of the test from free to cued recall, there was still no difference in test performance between the varied- and constant-encoding conditions, $t(33) = 1.024$, $p = .31$, $d = 0.18$, with the null hypothesis being 3.36 times more likely than the alternative hypothesis. In Experiment 3b, however, a different pattern emerged. When study words were presented three times, varied encoding produced an improvement in cued-recall performance as compared to constant encoding, contrary to the previous results. This difference was statistically significant, $t(44) = 3.524$, $p = .001$, $d = 0.53$, and the alternative hypothesis was 29.35 times more likely than the null.

The present results provide the first suggestion that variable encoding operating at the level of the same component across study presentations can yield benefits for memory performance. They are consistent with the notion that the match between the components tapped at encoding and retrieval is key for promoting the occurrence of encoding variability benefits. However, this pattern of benefits of encoding variability conditions emerged only with three presentations of study words in Experiment 3b. It would seem that the manipulation of encoding variability as implemented here is not always potent enough to benefit memory performance. Given that this result might be considered a little surprising, the next experiment is a conceptual replication of Experiment 3b.

Experiment 4

Experiment 3b was the first in the present series to document reliable effects of encoding variability on memory performance. The patterns documented so far suggest that in order to be able to observe benefits of encoding variability when this variability is introduced while keeping type of processing constant across repetitions, one needs both a strong encoding manipulation – exemplified here by at least three study cycles – and an appropriate test that taps into specific components of memory traces that are elaborated upon across variable study trials. Experiment 4 tested these contentions by conceptually replicating Experiment 3b. For Experiment 4, we modified both the materials and the specific memory test administered to participants in order to ensure the generalizability of the findings presented so far. We used the general outline of Experiment 3b but in conjunction with study words derived from associative norms rather than semantic categories, and with a different extra-list cued final memory test: associate-cued recall (e.g., Nelson, Kitto, Galea, McEvoy, & Bruza, 2013).

Method

Participants. Thirty-two students from Cardiff University (age range: 18-20, three males) took part in this experiment in exchange for course credit or payment.

Materials, Design, and Procedure. A new set of 44 study words and corresponding extra-list cues was compiled from the University of South Florida association norms (Nelson, McEvoy, & Schreiber, 1998). Associative extra-list cues were words which were moderately associated with study words (e.g., for the studied word 'neck' the extra-list cue used was 'giraffe'). As in Experiments 2b and 3b, the words were then split into two lists of 22 words, each of which was assigned to a separate study-test block. The list of questions used as orienting tasks was the same as in Experiment 3b. The design was the same as in previous experiments. The procedure was modelled on that from Experiment 3b, with three study presentations of each word and a cued-recall test, the only exception being the type of cues used at test. Instead of category names, associative extra-list cues were presented, and the participants' task was to type in next to each cue the studied word that was associated with it.

Results and Discussion

Descriptive statistics are presented in Table 1. Recall was scored as correct whenever a target word was produced, independent of whether it was provided next to the cue that it was associated with or next to a different cue. As in Experiment 3b, cued-recall performance was better in the varied- rather than in the constant-encoding condition, $t(31) = 2.824$, $p = .008$, $d = 0.50$, with a Bayes factor of 5.21 supporting the alternative hypothesis over the null.

These results conceptually replicate the findings in Experiment 3b. With three presentations of the study items, performing a different orienting task on each repetition confers benefits to memory performance as compared to performing the same orienting task on each repetition. These benefits were detected in the associate-cued recall test, just as they were detected with category cues in Experiment 3b. This contrasts with the null results obtained when free-recall performance was examined in Experiments 1 and 2. Together, these results indicate that, in agreement with the ideas developed by Glenberg (1979), variable encoding under different orienting tasks allows for establishing rich memory traces which can augment memory performance in tests that are sensitive to the elaboration of descriptive components of memory representations.

However, it is worth noting that so far our comparison of results across free- and cued-recall performance necessarily rests on comparisons between different experiments. To enable drawing stronger conclusions regarding any potential differences between the effects of encoding variability when cued- versus free-recall tests are considered, we conducted a single experiment contrasting both test types. For this reason, Experiment 5 combined the procedures of Experiments 2b and 3b in a single design, with test type being manipulated between participants.

Experiment 5

The present experiment contrasted directly cued- and free-recall tests within the encoding variability paradigm. We again used individual orienting questions targeting the descriptive component of memory representations in the procedure with three study cycles, manipulating constant versus variable encoding. We then administered different tests to two groups of participants, with one group completing a category-cued recall test (as in Experiment 3b) and one completing a free-recall test (as in Experiment 2b). Following the component-levels theory (Glenberg, 1979) and the results described so far, we predicted that the benefits of encoding variability would emerge in the extra-list cued-recall task – tapping descriptive components of memory representations – but not in the free-recall task – tapping predominantly structural and contextual components.

Method

Participants. Sixty-four students and graduates of Polish universities (age range: 18-57; 16 males), all of whom were fluent speakers of Polish, took part in this experiment for course credit or monetary compensation. Thirty-two participants were tested in each of the two experimental groups. The first 48 participants were tested in the laboratory, as in previous experiments. Due to the COVID-19 pandemic, the testing protocol was somewhat altered for the remaining 16 participants. To keep the testing conditions as close as possible to standard laboratory ones, those participants downloaded a standalone version of the experimental program to their own computers and completed the procedure while being under constant online supervision of the experimenter via a direct video link.[3]

Materials, Design, and Procedure. For the purpose of this and the subsequent experiments, all materials were translated into Polish. The number of words was increased to 54, each taken from a different semantic category. We also added new orienting questions, bringing the total to 108. Experiment 5 included two independent variables. Test type was manipulated between participants. The *free-recall* group went through the same procedure as participants in Experiment 2b, and the procedure in the *cued-recall* group was taken from Experiment 3b, with category cues being used at test. Encoding condition (varied versus constant) was manipulated within participants, as in all other

[3] We found no systematic differences between the results obtained in the laboratory and via the video link.

experiments. The experiment used unique orienting questions for each studied word. A schematic design of this experiment can be seen in Figure 2. Other than the language change, there were two other changes to the procedure. First, we increased the number of study lists from two to three (with 18 words per list, 27 total in the varied and 27 in the constant encoding condition) with the purpose of reducing error variance in the results. Second, for the 16 participants tested via a video link consent was collected as part of the experimental procedure rather than on a separate form before the start of the experiment.

Results and Discussion

Descriptive statistics can be found in Table 1. A repeated-measures 2 (test type: free recall, cued recall) x 2 (encoding condition: varied, constant) ANOVA revealed a significant main effect of test type, $F(1, 62) = 56.04$, $MSE = 1.34$, $p < .001$, $\eta_p^2 = 0.475$, with overall performance being higher in the cued-recall ($M = .74$) than in the free-recall group ($M = .53$). The main effect of encoding condition was not significant, $F(1, 62) = 1.87$, $MSE = 0.019$, $p = .176$, $\eta_p^2 = 0.029$, but the interaction between test type and encoding condition was, $F(1, 62) = 7.45$, $MSE = 0.076$, $p = .008$, $\eta_p^2 = 0.107$. In the cued-recall group, performance was higher when encoding was varied rather than constant, $t(31) = 2.83$, $p = .008$, $d = 0.50$, replicating the findings from Experiment 3b and – conceptually – Experiment 4, while in the free-recall group there was no significant difference between the two encoding conditions, $t < 1$, $d = 0.18$, with the numerical trend actually favoring the constant encoding condition.

An additional repeated-measures Bayesian ANOVA was further applied to the same data. The full results can be found in Table 2. The model that fared best included both main effects and the interaction, and there was anecdotal (main effect of test type), moderate (both main effects) to extreme (all other models) evidence (Lee & Wagenmakers, 2013) against the remaining models providing a better fit to the data. An analysis of effects with matched models confirmed that there was moderate evidence for including the interaction of test type and encoding condition ($BF_{inclusion}$ = 5.49). There was also extreme evidence for including test type in the model ($BF_{inclusion}$ = 27,850,000), while the evidence for not including the factor of encoding condition was anecdotal ($BF_{inclusion}$ = 0.41).

These results provide a clear replication of patterns found in previous experiments. They confirm in a single design that the effects of encoding variability on memory performance depend very much on the type of the test. While for extra-list cued-recall tests varied encoding consistently boosted performance compared to constant encoding – at least when the number of study repetitions was appropriate – this pattern also consistently failed to emerge when free-recall tests are used to measure performance. We argue that these results are consistent with the component-levels theory of Glenberg (1979) and demonstrate that variable-tasks encoding benefits memory retention as long as there is a match between the component enriched by variable encoding and the component tapped by a particular memory test.

Although the results obtained in Experiment 5 were consistent with our predictions and thus support the theories under test here, it needs to be acknowledged that a comparison between cued recall and free recall is far from ideal. This is because of limited experimental control over participants' strategies and retrieval processes in a free-recall task. While we argue that the null result in free recall reflects the test's reliance on contextual and structural components that occlude the role of the descriptive component enriched by our manipulation of encoding variability, alternative accounts can easily be proposed. For example, one could further develop an argument we have already mentioned in relation to the results of Experiments 2a and 2b (see Footnote 2) and stipulate that, in a free-recall task, participants engage in self-cueing by trying to mentally reinstate orienting questions in order to aid retrieval. For Experiments 1a and 1b, which used only two or three questions per list, this strategy would not yield any differences across varied- and constant-encoding conditions, as questions used in these conditions were presented equally often. However, in these experiments the effects of encoding variability could have been absent due to an overlap in features encoded in the variable encoding condition, consistent with the memory-as-discrimination framework (Nairne, 2002) and the concept of feature generality described by Glenberg (1979). On the other hand, for subsequent experiments, in which individual questions were used, it is straightforward to assume that self-cueing would be more effective in the constant-encoding condition, in which each question was repeated, than in the varied-encoding condition, in which each question was used once only. It can then be assumed that more effective self-cueing would bring free-recall performance up in the constant-encoding condition, while semantic elaboration would do the same in the varied-encoding condition, the outcome being a net null result in free recall. If so, this would be quite independent of the issue of the components at which variability was manipulated and which were tapped at retrieval.

Given the ambiguity associated with measuring memory processes via a free-recall task, it seemed necessary to provide additional evidence to strengthen our theoretical claims regarding the role of a match between components of memory representations elaborated on at encoding and those tapped by the final test. This was the purpose of Experiment 6.

Experiment 6

In the present experiment, we again examined the role of the final test in producing the effects of encoding variability. We preserved the extra-list cued-recall structure of the final test – one that revealed benefits of encoding variability in Experiments 3b, 4, and 5 – but used a test that would tap features other than those encoded by responding to our orienting questions, resulting in the abolishment of the benefits of encoding variability. In Experiment 6, we used a rhyme-cued test for this purpose.

All experiments presented in this study used orienting questions pertaining to semantic aspects of the to-be-remembered items: size, overall appearance, cost, function, and other similar attributes. However, it is important to note that study materials are typically endowed with other features as well, such as phonology, orthography (for words), or perceptual aspects. Indeed, Glenberg (1979) has argued that features encoded at the descriptive level depend on the particular orienting task, using an example of deep encoding resulting in semantic features being underscored and shallow encoding resulting in the elaboration of perceptual/orthographic features. Although one can treat different features described here as different manifestations of the descriptive component, it could also be useful to describe them as subcomponents in the hierarchy of levels proposed by Glenberg. The benefit of treating different facets of memory representations as descriptive subcomponents is that the same principle of encoding-retrieval match, which lies at the core of the component-levels theory, could then be applied to this lower level as well. Thus, if variable encoding leads to enriching memory representations with semantic features, then this should be detected in memory tests tapping semantic features, but not phonemic or orthographic ones (see also McDaniel & Masson, 1985, for the examination of encoding variability at semantic and phonemic levels). Note that this extension of Glenberg's theory at the descriptive level would bring it close to the transfer-appropriate processing framework (Morris, Bransford, & Franks, 1977), which also assigns a central role to the encoding-retrieval match in shaping patterns of memory retention. Following this additional specification of the descriptive component, here we focus on phonology as a specific aspect of the studied words and argue that given that our encoding manipulation targets semantic features, it should not affect performance in a rhyme-cued test.

Method

Participants. Thirty-five students (age range: 21-56, five males) who were fluent speakers of Polish took part in this experiment in exchange for course credits. All students were tested via a video link, with constant supervision of the experimenter.

Materials, Design, and Procedure. The materials were the same as in Experiment 5. The design procedure closely followed that from Experiments 3b and 4, with three cycles of learning and unique orienting questions being used for each studied word, and only the encoding condition being manipulated within participants. As in Experiment 5, participants studied three separate lists of words. The only change to the procedure was introduced at test. For each studied word, a unique rhyming cue was chosen (e.g., if participants studied TABLE, it could be cued by ENABLE). These cues were displayed one at a time and participants were instructed to match a studied word to each of the rhyming cues.

Results and Discussion

Descriptive statistics are presented in Table 1. There was no significant difference in recall performance across varied- and constant-encoding conditions, $t < 1$, $d = 0.07$, and a Bayesian paired-samples *t*-test revealed that the null hypothesis was 5.05 times as likely as the alternative hypothesis. The Bayesian test is of particular importance here as it provides positive evidence for the lack of benefits of encoding variability – an effect directly in contrast to those observed with category- (Experiments 3b and 5) and associate-cued recall (Experiment 4). The results thus clearly demonstrate that it is not the particular format of a final test that is responsible for the patterns of positive and null effects of encoding variability discussed here. Even in a cued-recall test a null effect can be obtained, as long as this test does not tap the same (sub)component of memory representations affected by the encoding variability manipulation. While positive effects of encoding variability emerged in tests that tapped semantic features, matching the features targeted by our orienting questions, they were absent from a test tapping phonemic features. Thus, the present results clearly establish that it is the interplay between encoding and testing conditions that is responsible for the consistent emergence of encoding variability benefits in our paradigm.

Experiment 7

When introducing Experiments 2a and 2b, we have argued that there are theoretical reasons to expect that the benefits of encoding variability may depend not only on the elaboration of individual memory traces due to the use of variable tasks across repetitions but also on the specificity of this elaboration. Building on the memory-as-discrimination framework (Nairne, 2002), we have hypothesized that one reason these benefits are so elusive is that many previous studies – including the study by Young and Bellezza (1982), but also our own Experiments 1a and 1b – manipulated encoding variability by using only a few different tasks, creating large overlap in terms of the semantic details encoded across different study items. For this reason, starting from Experiment 2a and 2b, we used unique questions to foster processing that would not only enrich the semantic (descriptive) components of memory representations but would also serve to differentiate the features encoded across different representations. However, as a result we still do not know whether the benefits of encoding variability obtain when only a few orienting tasks are used. Although Experiments 1a and 1b which used non-unique questions failed to document the benefits of encoding variability, these experiments employed free-recall tests, which we now know to be insufficiently sensitive to the effects of encoding variability with variable tasks. We thus conducted Experiment 7 to directly compare the effectiveness of unique and non-unique orienting tasks in promoting the benefits of encoding variability.

This final experiment allowed us also to address one methodological limitation that arose for all previous experiments reported here. The general procedure used so far required participants to answer orienting questions at encoding and the study words were displayed for as long as it took to provide an answer. This feature of our methodology meant that we had no control over presentation times. Under these conditions, it is possible that our two conditions of interest – constant and varied encoding – differed in the amount of time participants spent studying the presented words. Indeed, there is good reason to expect that with repeated questions in the constant encoding conditions, participants would speed up responding as compared to when novel questions were asked for study words in the variable encoding condition.[4] This difference in study times across conditions could contribute to differences in memory performance. Consequently, in Experiment 7 we modified our procedure in such a way that all study words were displayed for a constant duration across all cycles of learning, independently of how long it took participants to respond to orienting questions.

Method

Participants. Seventy-seven students and graduates (age range: 18-61, 21 males) took part in this experiment in exchange for course credit or payment. All participants were fluent speakers of Polish and were tested in laboratory conditions. This experiment was conducted before Experiments 5 and 6, and is presented last for the ease of exposition. The sample size was this time larger than in the previous experiments because of the addition of another factor of question type, but it was again constrained by the requirement to obtain informative Bayes factors.

---

[4] The way we programmed our experiments meant that we had relevant data only for experiments using three study cycles. When we looked at these results, the intuition that participants would spend progressively less time on responding to repeated questions in the constant encoding conditions was confirmed (see Supplemental materials).

Materials, Design, and Procedure. Participants were presented with two lists of 18 words to learn, each list studied in three cycles. Whereas in all previous experiments employing three cycles of learning the two experimental study-test blocks were the same except for the change of materials, in the present experiment one block was assigned to the *unique questions condition* whereas the other block was assigned to the *overlapping questions condition*. In the unique questions condition, participants were asked individual questions for each of the study words, which changed across repetitions for words in the varied encoding condition but remained the same for words in the constant encoding condition (as in Experiments 2b, and 3b-6). In the overlapping questions condition, only three different questions were used and each question was asked for one third of the study list (as in Experiment 1b). Here, questions for the study words from the constant encoding condition also remained unchanged across repetitions whereas questions for the study words from the variable encoding conditions were switched across repetitions so that each word was studied once with each of the three questions across the study cycles. The design of the experiment was thus 2 (unique vs. overlapping questions) x 2 (constant vs. variable encoding), with the first factor manipulated across study blocks and the second factor manipulated within blocks. The order of blocks was counterbalanced across participants. The details of the procedure were modelled on those from Experiment 3b, with the only exception being the fixed presentation time at study: this time all words were presented for 2 seconds during study (based on reaction times from Experiment 1b) and participants were asked to provide their response to the orienting questions within this time window.

Results and Discussion

Descriptive statistics are presented in Table 1. A 2 (question: unique, overlapping) x 2 (encoding condition: varied, constant) repeated-measures ANOVA revealed only a significant main effect of encoding condition, $F(1, 76) = 16.67$, $MSE = 0.41$, $p < .001$, $\eta_p^2 = 0.180$, with higher performance in the varied encoding condition ($M = .68$) than in the constant encoding condition ($M = .61$). Neither the main effect of question, nor the interaction was significant, $F < 1$, $\eta_p^2 = .009$, and $F(1, 76) = 1.21$, $MSE = 0.03$, $p = .275$, $\eta_p^2 = 0.016$, respectively.

Additionally, we conducted a Bayesian repeated-measures ANOVA to further quantify the evidence for the effects of interest. The results are presented in Table 3. Out of all competing models, the model including only the encoding condition fared best, and there was moderate to extreme evidence against any of the other models providing a better fit to the data compared to the encoding condition only model. An analysis of effects with matched models confirmed that there was very strong evidence for including the encoding condition in the model ($BF_{inclusion}$ = 96.90), but not the main effect of question ($BF_{inclusion}$ = 0.18) nor the interaction ($BF_{inclusion}$ = 0.28).

The present experiment contributes in several ways to our general conclusions. First, we once again replicated the reliable benefits of encoding variability with a category-cued recall task, which speaks to the robustness of the effect – something of note given the widely contradictory patterns of data in the published literature and also our unsuccessful efforts to find those benefits with a free-recall task in Experiments 1, 2, and 5, and a rhyme-cued recall task in Experiment 6. Second, this replication shows that the discussed effect cannot be assigned to the difference in study times across constant and variable encoding conditions. Third, we now show that, contrary to our initial assumption, the benefits of encoding variability do not seem to depend strongly on the uniqueness of features embedded in memory representations of different study items. It is worth noting, however, that the effect size associated with benefits of encoding variability with unique questions (d = 0.43) was larger in the present experiment that the corresponding effect size associated with these benefits when overlapping questions were used (d = 0.23). Thus, although the current results certainly do not substantiate any theoretical claims regarding the role of item differentiation in yielding benefits of encoding variability, we would still suggest that, from a practical perspective, researchers wishing to investigate encoding variability with varied tasks should consider making their encoding tasks unique.

General Discussion

This series of experiments was devoted to the issue of encoding variability. It has been long argued that varying the way in which information is encoded during study should enhance the richness of memory traces, conferring benefits for memory performance due to the abundance of routes by which rich memory traces can be retrieved (e.g., Glenberg, 1979; Melton, 1970). The principle of encoding variability has been embedded in theoretical models of memory and has been used as an explanatory mechanism for some fundamental patterns of memory performance, such as the spacing effect (see Benjamin & Tullis, 2010; Maddox, 2016, for recent discussions). Yet empirical support for the notion of encoding variability has been underwhelming, motivating us to systematically pursue this phenomenon. Here we demonstrate that encoding variability does indeed reliably improve memory performance when specific conditions are met. In particular, the benefits of encoding variability for memory performance were revealed when a sufficient number of study presentations – three in our case – was used in the experimental procedure, and when a specific type of test was used to assess memory. Specifically, while benefits of encoding variability were revealed in extra-list cued-recall tests that tapped semantic aspects of the studied words, they were consistently absent when memory was assessed via free recall, and also when a cued-recall test required access to phonemic rather than semantic information.

We start our discussion by considering boundary conditions for our effects of encoding variability. First, the benefits of encoding variability were revealed when three study cycles were used but – in Experiment 3a – we failed to observe these benefits under conditions that were identical apart from the fact that only two study cycles were used. This is currently an isolated finding in need of replication, but it does suggest that encoding variability benefits might be less likely to be found when opportunities for additional elaboration in the variable-encoding condition are limited. This seems intuitive because more study opportunities in the varied-encoding condition mean more ways in which memory representation is elaborated. In this regard, it is worth noting the recent contextual variability results by Smith and Handy (2016). Participants studied Tagalog-English pairs superimposed over context photographs and later completed between one and five cycles of retrieval practice (with feedback) for those pairs with either original or varied contexts. On a final cued-recall test, the benefits of contextual variability occurred only after completing four or five retrieval trials. Our results and those of Smith and Handy point to a possible parallel between encoding variability imposed by orienting tasks and by changing contextual features.

Second, the benefits of encoding variability in our study emerged only when extra-list cued-recall tests targeting semantic information were used. This was directly tested in Experiment 5, which contrasted cued- and free-recall testing conditions in a single design, but the same pattern can also be seen across experiments employing three item presentations, with all cued-recall testing conditions (barring the non-semantic testing condition of Experiment 6) and none of the free-recall conditions showing the benefits of encoding variability. To further strengthen this claim, we conducted a combined Bayesian ANOVA on the results of experiments and conditions which used three study presentations, unique questions, and either a free- or a semantic cued-recall test. This analysis thus was conducted on data from Experiments 2b, 3b, 4, 5, and the unique-questions condition from Experiment 7, and included factors of encoding condition (varied, constant), test type (cued recall, free recall), as well as experiment. There was extreme evidence for the interaction of encoding condition and test type, $BF_{(inclusion)} = 425.28$, and moderate evidence against a triple interaction of encoding condition, test type, and experiment, $BF_{(inclusion)} = 0.15$. Also, evidence for the benefits of encoding variability on cued recall – collapsed across experiments – was extreme, with $BF_{(inclusion)} = 5{,}862{,}000$, while the evidence against the same effect in free recall was ambiguous, with $BF_{(inclusion)} = 0.44$. The results of this combined analysis thus further confirm that, consistently across our set of experiments, cued-recall testing reliably produced benefits of encoding variability, while free-recall testing did not.

It is worth underscoring that the pattern of encoding variability benefits emerged in our study every single time when semantic-dependent cued-recall tests were used (coupled with three study presentations), regardless of whether they employed category (Experiments 3b, 5, and 7) or associated cues (Experiment 4). This is an important observation inasmuch as many previous studies of encoding variability, including the most recent investigation by Huff and Bodner (2014), focused mostly on free recall. We argue that the issue of why the benefits of encoding variable are observed only under particular testing conditions can be best addressed by considering again the component-levels theory proposed by Glenberg (1979) to describe the contribution of encoding variability to the effects of spaced practice. Glenberg argued that different memory tests may be differentially sensitive to different components of memory representations. From this perspective, the manipulation of encoding variability that we used – varying orienting questions that referred to conceptual aspects of the study words – served to change the semantic features encoded for study words across constant and variable encoding conditions, with greater semantic richness of representations established under the variable encoding conditions. Importantly, all features added to memory representations related to the same semantic component, thus holding type of processing constant across repetitions. This greater richness or elaboration of representations established as a result of variable-task encoding translated into better performance when semantic features were queried at test because with a greater number of features encoded in the variable encoding condition there is a greater chance that semantic features embedded in a retrieval cue match the contents of memory. The extra-list cued-recall tests, specifically designed to tap into memory for items with the exclusion of contextual (Hanczakowski & Mazzoni, 2013) and structural components (Anderson, 2003; Zawadzka & Hanczakowski, 2019), proved sensitive to the additional semantic elaboration conferred by variable encoding. However, a free-recall task, which depends largely on cueing with contextual information and other retrieved items was not sensitive to the same benefits, which also failed to emerge in a test using extra-list cues targeting phonemic rather than semantic information (Experiment 6).

Although the present study was designed mostly to confirm that benefits of encoding variability can be reliably observed, it can also serve as a stepping-stone for further explorations of this fundamental mechanism of repeated learning. The first issue that remains to be resolved is how the present results inform the most recent framework of encoding variability proposed by Huff and Bodner (2014). Our work was concerned with conditions of task variability only and no effort was made to test a hypothesis according to which the benefits of encoding variability are more likely to emerge when variability in terms of processes – different components elaborated across repetitions – is instantiated. What our study contributes here is a proper baseline against which any potential benefits of processing variability could be assessed. While Huff and Bodner documented benefits of processing variability, varying both structural and descriptive components across study repetitions, against the baseline of both constant encoding (in terms of both processes and tasks) and task variability, our study clearly demonstrates that these two levels of comparison need not be equal. Further studies should directly contrast performance under variable processing conditions against the baseline of conditions under which task variability by itself leads to better performance than constant encoding.

Another issue is highlighted by the consideration of the role of encoding-retrieval match in revealing the benefits of encoding variability. In the present study we varied test conditions across experiments, but we consistently used the same manipulation at encoding, targeting the descriptive, or more specifically semantic, component of memory representations. This focus on a single component results in the observed patterns of single dissociations, where performance on tests tapping the semantic component is affected by variable encoding, while performance relying on other components is not. However, if the reasoning embedded in the component-levels theory is correct, it should be possible to target – via encoding variability manipulations – other components, leading to patterns of dissociations across tests tapping these components. The logic here would be very similar to the one often proposed within the transfer-appropriate processing framework (Morris et al., 1977), where manipulations of perceptual features are thought to exert influence on perceptual but not conceptual tests, while manipulations of conceptual features are thought to exert influence on conceptual but not perceptual tests (Blaxton, 1989). The logic of the component-levels theory is essentially the same as the logic of the transfer-appropriate processing framework, with an addition of a clear specification of components of memory representations that may determine whether encoding and retrieval conditions are matched. Interestingly, a study on the testing effect by Veltre, Cho, and Neely (2014), with retrieval practice rather than restudy used for revision of original materials, found support for the role of the match of features promoted by retrieval practice and required on the final memory test in determining memory performance, with evidence of a double dissociation across semantic and orthographic features. Such a dissociation is yet to be directly demonstrated with encoding variability procedures – a potential way of merging the theory described here with the transfer-appropriate processing framework.

As mentioned in the Introduction, the concept of encoding variability has been most often used to account for the benefits of spacing, which is sometimes assigned to increased variability in terms of the contextual component when a temporal delay is inserted between study presentations. The present study provides support for the overall mechanism by which encoding variability is thought to operate, albeit at the level of the descriptive rather than contextual component. However, it is important to note that recent theorizing about the spacing effect indicates that encoding variability *per se* is insufficient to account for the benefits of spacing and an additional mechanism, operating alongside encoding variability (see Siegel & Kahana, 2014), is implicated, usually identified as study-phase retrieval, or – to give this process its alternative name – reminding (Benjamin & Tullis, 2010; Greene, 1989; Hintzman, 2004; Maddox, Pyc, Kauffman, Gatewood, & Schonhoff, 2018). Specifically, it is often argued that benefits of spacing arise because delaying repetitions creates an opportunity for spontaneous retrieval of previous study opportunities when a given item is again presented and such retrieval subsequently benefits memory performance in a way akin to the testing effect (Roediger & Karpicke, 2006). Acknowledging the role of reminding in repeated learning begs the question of its potential role in the present study.

In our own recent investigation into the issue of reminding (Zawadzka, Simkiss, & Hanczakowski, 2018), we have shown that the likelihood of being reminded of the previous occurrence of a study item is increased when this item is repeatedly studied in the same context. This observation joins other studies which also suggest that the likelihood of being reminded of the previous study opportunity is a function of the similarity of conditions across repetitions (Verkoeijen, Rikers, & Schmidt, 2004; Tullis, Braverman, Ross, & Benjamin, 2014). If so, then one could argue that the conditions of varied encoding are precisely those that minimize the chances of being reminded of previous study opportunities, thus depriving encoded representations of additional benefits stemming from successful retrieval. Applying this logic to our study, it stands to reason that orienting questions served as semantic contexts, possibly triggering reminding whenever these questions were repeated in the constant encoding condition, thus benefiting memory retention. Moreover, this contribution of reminding under constant encoding conditions is potentially capable of accounting for the previously mentioned results of Young and Bellezza (1982), who also used orienting tasks as their manipulation of encoding variability and found better performance with the constant rather than varied encoding tasks.[5] If one assumes that reminding is an important factor in repeated-study procedures, then it can be argued that the benefits of encoding variability that we

---

[5] Interestingly, Young and Bellezza (1982, Experiment 4) used two cycles of learning, overlapping orienting tasks and a free recall test – all conditions that can potentially undermine the benefits of variable encoding.

document here were indeed potent enough to actually overcome the benefits of reminding that selectively enhanced performance in the constant condition.

Given the theoretical focus of our study, it is interesting to ask how reminding chimes with the component-levels theory of Glenberg (1979). By arguing that repeated presentations under varied-encoding conditions serve to enrich memory representation by imbuing them with additional features assembled in various components of memory representations, this theory essentially describes the process of updating the *same* memory representations across repetitions. It thus implies that reminding occurs *always* when items are repeated. In other words, the component-levels theory, as originally formulated by Glenberg, assumes a 100% likelihood of reminding. A recent surge of studies on reminding clearly suggests that this is an oversimplification and that various factors affect the likelihood of reminding, among others contextual match across presentations (Zawadzka et al., 2018), attentional resources at repetition (Sahakyan & Malmberg, 2018), time to encode repeated stimuli (Negley, Kelley, & Jacoby, 2018). If reminding serves to occlude the benefits of encoding variability and the likelihood of reminding can be systematically varied, then it is possible that the likelihood of detecting the benefits of encoding variability will also depend on the same factors. It is perhaps worth returning here to one aspect of our data: the perplexing lack of encoding variability benefits with two study presentations in Experiment 3a. Although we have already discussed this pattern from the perspective of a sheer strength of the experimental manipulation, it is also possible to look at this problem from the reminding perspective. It is possible that additional repetitions generally augment the likelihood of reminding, also in the variable encoding condition, creating a situation in which reminding occurs for virtually all studied items. If so, this would largely equate the workings of this mechanism across varied- and constant-encoding conditions. If this stipulation is correct, then this would mean that multiple presentations of stimuli during study create an environment similar to the idealized situation from the component-levels theory, where reminding is maximized across conditions and thus its impact on the results is minimized, making it possible for the benefits of encoding variability to be clearly outlined.

At this point, the possible interplay of encoding variability and reminding remains a speculation. Recent studies on reminding were not much concerned with encoding variability – an idea that has not been extensively tested in recent years – and our own investigation was not concerned with reminding. We had no measure of reminding in our procedure and thus we cannot confirm directly whether reminding was more prevalent in the constant rather than varied encoding condition or whether it was indeed occurring equally often in both conditions. Also, a detailed examination of the issue of reminding when study items are repeated is hindered by the fact that reminding is a theoretical construct the experimenters do not usually observe but merely infer from the performance patterns observed in subsequent memory tests. To remedy this problem, we would argue that studies on repeated encoding would benefit from incorporating such direct measures of reminding, either in the form of explicit reports (e.g., Wahlheim & Jacoby, 2013; McKinley & Benjamin, 2020) or in the form of measures collected during encoding that remain sensitive to the effects of reminding (e.g., judgments of learning, as investigated by Tullis et al., 2014; Zawadzka et al., 2018, or study times, as demonstrated by McKinley, Ross, & Benjamin, 2019).

Conclusion

We found what we consider compelling evidence that variable encoding is capable of augmenting memory performance. Given the highly inconsistent findings on encoding variability reported to date, it perhaps comes as no surprise that specific conditions need to be met to reliably observe the benefits of encoding variability. The role of the encoding-retrieval match, highlighted by the component-levels theory of Glenberg (1979), comes to the fore here. While varied encoding is necessarily applied at study, only tests that tap features encoded as a result of varying study conditions are capable of revealing the memory benefits such encoding confers. The present study thus demonstrates that encoding variability can potentially serve as an overarching principle of effective encoding across various components of memory representations, but careful ways of assessing its benefits need to be employed in experimental studies. We believe that further investigation of the encoding variability principle should be of high priority for the sake of memory models that explicitly invoke the concept of contextual drift and the associated benefits of variable encoding, as well as for more application-oriented research conducted in educational settings.

References

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

Beaman, C. P., Hanczakowski, M., Hodgetts, H. M., Marsh, J. E., & Jones, D. M. (2013). Memory as discrimination: What distraction reveals. *Memory & Cognition, 41,* 1238-1251.

Bellezza, F. S., & Young, D. R. (1989). Chunking of repeated events in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 990-997.

Bird, C. P., Nicholson, A. J., & Ringer, S. (1978). Resistance of the spacing effect to variations in encoding. *American Journal of Psychology, 91,* 713-721.

Blaxton, T. A. (1989). Investigating dissociations among memory measures: Support for a transfer-appropriate processing framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 657-668.

Bradshaw, G. L., & Anderson, J. R. (1982). Elaborative encoding as an explanation of levels of processing. *Journal of Verbal Learning and Verbal Behavior, 21,* 165-174.

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.

Delaney, P. F., Verkoeijen, P. P. J. L., & Spirgel, A. (2010). Spacing and testing effects: A deeply critical, lengthy, and at times discursive review of the literature. In B. H. Ross (Ed.), *Psychology of Learning and Motivation, Vol. 53* (pp. 63-148). Burlington: Academic Press.

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.

Estes, W. K. (1955). Statistical theory of distributional phenomena in learning. *Psychological Review, 62,* 369-377.

Gallo, D. A., Meadow, N. G., Johnson, E. L., & Foster, K. T. (2008). Deep levels of processing elicit a distinctiveness heuristic: Evidence from the criterial recollection task. *Journal of Memory and Language, 58,* 1095-1111.

Glanzer, M., & Duarte, A. (1971). Repetition between and within languages in free recall. *Journal of Verbal Learning and Verbal Behavior, 10,* 625-630.

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

Goh, W. D., & Lu, S. H. (2012). Testing the myth of the encoding-retrieval match. *Memory & Cognition, 40,* 28-39.

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

Hanczakowski, M., & Mazzoni, G. (2011). Both differences in encoding processes and monitoring at retrieval reduce false alarms when distinctive information is studied. *Memory, 19,* 280-289.

Hanczakowski, M., & Mazzoni, G. (2013). Contextual match and cue-independence of retrieval-induced forgetting. Testing the prediction of the model by Norman, Newman, and Detre (2007). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39,* 953-958.

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

Hintzman, D. L., & Stern, L. D. (1978). Contextual variability and memory for frequency. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 539-549.

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology, 46,* 269-299.

Huff, M. J., & Bodner, G. E. (2014). All variations of encoding variability are not created equal: Separating variable processing from variable tasks. *Journal of Memory and Language, 73,* 43-58.

Huff, M. J., Bodner, G. E., & Fawcett, J. M. (2015). Effects of distinctive encoding on correct and false memory: A meta-analytic review of costs and benefits and their origins in the DRM paradigm. *Psychonomic Bulletin & Review, 22,* 349-365.

Hunt, R. R., & Einstein, G. O. (1981). Relational and item-specific information in memory. *Journal of Verbal Learning and Verbal Behavior, 20,* 497-514.

JASP Team (2018). JASP(Version 0.8.4) [Computer software]

Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review, 25,* 178-206.

Lee M. D., & Wagenmakers E. J. (2013). *Bayesian Cognitive Modeling: A Practical Course*. Cambridge, UK: Cambridge University Press.

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

Maddox, G. B., Pyc, M. A., Kauffman, Z. S., Gatewood, J. D., & Schonhoff, A. M. (2018). Examining the contributions of desirable difficulty and reminding to the spacing effect. *Memory & Cognition, 46,* 1376-1388.

Maskarinec, A. S., & Thompson, C. P. (1976). The within-list distributed practice effect: Tests of the varied context and varied encoding hypotheses. *Memory & Cognition, 4,* 741-746.

McDaniel, M. A., & Masson, M. E. J. (1985). Altering memory representations through retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 371-385.

McKinley, G. L., & Benjamin, A. S. (2020). The role of retrieval during study: Evidence for reminding from overt rehearsal. *Journal of Memory and Language, 114,* 104128.

McKinley, G. L., Ross, B. H., & Benjamin, A. S. (2019). The role of retrieval during study: Evidence of reminding from self-paced study time. *Memory & Cognition, 47,* 877-892.

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.

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

Nairne, J. S. (2002). The myth of the encoding-retrieval match. *Memory, 10,* 389-395.

Negley, J. H., Kelley, C. M., & Jacoby, L. L. (2018). The importance of time to think back: The role of reminding in retroactive effects of memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44,* 1352-1364.

Nelson, D. L., Kitto, K., Galea, D., McEvoy, C. L., & Bruza, P. D. (2013). How activation, entanglement, and searching a semantic network contribute to event memory. *Memory & Cognition, 41,* 797-819

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

Postman, L., & Knecht, K. (1983). Encoding variability and retention. *Journal of Verbal Learning & Verbal Behavior, 22,* 133-152.

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

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88,* 93-134.

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17,* 249-255.

Rose, R. J. (1984). Processing time for repetitions and the spacing effect. *Canadian Journal of Experimental Psychology, 83,* 527-550.

Sahakyan, L., & Malmberg, K. J. (2018). Divided attention during encoding causes separate memory traces to be encoded for repeated events. *Journal of Memory and Language, 101,* 153-161.

Siegel, L. L., & Kahana, M. J. (2014). A retrieved context account of spacing and repetition effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40,* 755-764.

Simmons, J. P., Nelson, L. D., & Simohnson, U. (2011). False-positive psychology. Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22,* 1359-1366.

Slamecka, N. J., & Barlow, W. (1979). The role of semantic and surface features in word repetition effects. *Journal of Verbal Learning and Verbal Behavior, 18,* 617-627.

Smith, S. M., Glenberg, A. M., & Bjork, R. A. (1978). Environmental context and human memory. *Memory & Cognition, 6,* 342-353.

Smith, S. M., & Handy, J. D. (2014). Effects of varied and constant environmental contexts on acquisition and retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40,* 1582-1593.

Smith, S. M., & Handy, J. D. (2016). The crutch of context-dependency: Effects of contextual support and constancy on acquisition and retention. *Memory, 24,* 1134-1141.

Tullis, J. G., Benjamin, A. S., & Ross, B. H. (2014). The reminding effect: Presentation of associates enhances memory for related words in a list. *Journal of Experimental Psychology: General, 143,* 1526-1540.

Tullis, J. G., Braverman, M., Ross, B. H., & Benjamin, A. S. (2014). Remindings influence the interpretation of ambiguous stimuli. *Psychonomic Bulletin & Review, 21,* 107-113.

van Overschelde, J. P., Rawson, K. A., & Dunlosky J. (2004). Category norms: An updated and expanded version of Battig and Montague (1969) norms. *Journal of Memory and Language, 50,* 289-335.

Veltre, M. T., Cho, K. W., & Neely, J. H. (2014). Transfer-appropriate processing in the testing effect. *Memory, 23,* 1229-1237.

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

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.

Wagenmakers, E. J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., … Morey, R. D. (2018). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review, 25,* 58-76.

Wahlheim, C. N., & Jacoby, L. L. (2013). Remembering change: The critical role of recursive remindings in proactive effects of memory. *Memory & Cognition, 41,* 1-15.

Watkins, O. C., & Watkins, M. J. (1975). Buildup of proactive inhibition: A cue overload effect. *Journal of Experimental Psychology: Human Learning and Memory, 1,* 442-453.

Weinstein, Y., Madan, C. R., & Sumeracki, M. A. (2018). Teaching the science of learning. *Cognitive Research: Principles and Implications, 3:2*.

Xue, G., Dong, Q., Chen, C., Lu, Z., Mumford, J. A., & Poldrack, R. A. (2010). Greater neural pattern similarity across repetitions is associated with better memory. *Science, 330,* 97-101.

Zawadzka, K., & Hanczakowski, M. (2019). Two routes to memory benefits of guessing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45,* 1748-1760.

Zawadzka, K., Simkiss, N., & Hanczakowski, M. (2018). Remind me of the context: Memory and metacognition at restudy. *Journal of Memory and Language, 101,* 1-17.

Table 1.

Means (Standard Deviations) for Recall Performance as a Function of Test Type, Question Type, and Encoding Condition

Number of Presentations and Experiment	Cued recall				Free recall			
	Unique questions		Overlapping questions		Unique questions		Overlapping questions	
	Varied encoding	Constant encoding	Varied encoding	Constant encoding	Varied encoding	Constant encoding	Varied encoding	Constant encoding
Two presentations								
Experiment 1a	-	-	-	-	-	-	.55 (.13)	.58 (.14)
Experiment 2a	-	-	-	-	.49 (.11)	.49 (.11)	-	-
Experiment 3a	.63 (.23)	.61 (.23)	-	-	-	-	-	-
Three presentations								
Experiment 1b	-	-	-	-	-	-	.64 (.13)	.65 (.13)
Experiment 2b	-	-	-	-	.45 (.14)	.48 (.15)	-	-
Experiment 3b	.75 (.14)	.67 (.16)	-	-	-	-	-	-
Experiment 4	.75 (.15)	.68 (.12)	-	-	-	-	-	-
Experiment 5	.77 (.09)	.70 (.16)	-	-	.52 (.13)	.54 (.13)	-	-
Experiment 6	.37 (.14)	.36 (.12)	-	-	-	-	-	-
Experiment 7	.68 (.22)	.59 (.25)	.68 (.23)	.63 (.24)	-	-	-	-

Table 2.

Bayesian Model Comparisons for Experiment 5 Data.

Model	P(M)	P(M|Data)	BF_M	BF_{10}	error%
test + encoding condition + test * encoding condition	0.20	0.614	6.371	1.000	
test	0.20	0.274	1.508	0.446	2.73
test + encoding condition	0.20	0.112	0.504	0.182	5.91
null	0.20	9.929e-9	3.972e-8	1.616e-8	2.36
encoding condition	0.20	3.921e-9	1.568e-8	6.383e-9	2.81

Note: P(M) shows uniform prior probabilities for all competing models. P(M|Data) shows posterior model probabilities. BF_M shows the change from prior to posterior odds. BF_{10} shows how much of an improvement compared to the best model (top row) each additional model provides.

Table 3.

Bayesian Model Comparisons for Experiment 7 Data.

Model	P(M)	P(M|Data)	BF_M	BF_{10}	error%
encoding condition	0.20	0.802	16.196	1.000	
question + encoding condition	0.20	0.147	0.691	0.184	2.40
question + encoding condition + question * encoding condition	0.20	0.041	0.171	0.051	2.29
null	0.20	0.008	0.034	0.010	1.27
question	0.20	0.001	0.006	0.002	1.51

Note: P(M) shows uniform prior probabilities for all competing models. P(M|Data) shows posterior model probabilities. BF_M shows the change from prior to posterior odds. BF_{10} shows how much of an improvement compared to the best model (top row) each additional model provides.