



Examining the role of context variability in memory for items and associations

William R. Aue^{1,2} · Jessica M. Fontaine¹ · Amy H. Criss¹

Published online: 30 April 2018
© Psychonomic Society, Inc. 2018

Abstract

What properties of a word make it easy or difficult to remember? Word frequency and context variability are separate, closely related word properties that have disparate influences on memorability. The influence of word frequency changes depending on the memory task, with high-frequency words tending to be recalled better and low-frequency words to be recognized better. Conversely, low-context-variability words tend to be remembered better across tasks. One proposed explanation for the low-variability advantage is that low-variability words are easier to associate with the experimental context, given that they are associated with fewer extra-experimental contexts. On the basis of this explanation, it has been suggested that the formation of interitem associations during encoding should interfere with the formation of item-to-context associations, attenuating the low-variability advantage. Across experiments, we tested whether focusing on interitem associations disrupted the low-variability advantage, by manipulating encoding tasks, test expectancy, final test condition, word frequency, and context variability. Focusing on interitem associations did not harm performance for low-variability words. Words low in both frequency and variability were recognized better, but word pairs composed of high-frequency, low-variability words were recognized better in associative recognition. On the basis of the data, we suggest that focusing on interitem associations does not come at the expense of item-to-context associations. Moreover, the data further support the idea that frequency and variability are distinct properties.

Keywords Memory · Recognition · Word recognition · Associative memory

What properties of words afford a memorability advantage? For example, if you are trying to remember items that you purchased from the store last week, are you more likely to remember that you purchased IODINE or a SHIRT? A variety of properties can contribute to memorability. Two examples of such properties are word frequency and context variability. *Word frequency* is a representation of how often a word tends to be encountered in the environment (e.g., Glanzer & Adams, 1985; Hemmer & Criss, 2013), whereas *context variability* represents the variety of different scenarios in which a word is encountered in the environment (Steyvers & Malmberg, 2003). If trying to recall the items in the example above, one possibility is that we would tend to recall SHIRT more often

than IODINE, because SHIRT is high-frequency and perhaps easier to generate. Another possibility is that we might tend to recall IODINE more often than SHIRT, because IODINE has low context variability, making it easier to associate to that *particular* shopping trip.

Word frequency and context variability influence memory in different ways, producing disparate patterns of data. The aim of this article is to expand our understanding of context variability by examining how changes to the testing scenario influence the phenomenon.

Word frequency

How word frequency impacts memory performance depends, in part, on how memory is tested (e.g., Glanzer & Bowles, 1976; Karlsen & Snodgrass, 2004). For example, during single-item recognition, participants are asked to distinguish previously studied words (i.e., targets) from unstudied words (i.e., foils). In this task, low-frequency words are remembered better. The low-frequency benefit comes in the form of a

✉ William R. Aue
william.aue@gmail.com

¹ Department of Psychology, Syracuse University, Syracuse, NY, USA

² Department of Psychological Sciences, Purdue University, West Lafayette, IN, USA

mirror effect, in which the studied targets are correctly endorsed more often (i.e., higher hit rate [HR]) and the unstudied foils are incorrectly endorsed less often (i.e., false alarm rate [FAR]), relative to high-frequency words (Balota & Neely, 1980; Chalmers, Humphreys, & Dennis, 1997; Criss & Shiffrin, 2004; Glanzer & Adams, 1985, 1990; Hemmer & Criss, 2013; Hintzman, Caulton, & Curran, 1994; Rao & Proctor, 1984; Stretch & Wixted, 1998; Wixted, 1992)—a pattern known as the *word frequency mirror effect* (Glanzer & Adams, 1985).¹ Forced choice measures of recognition support the idea that the low-frequency benefit is related to accuracy rather than bias (Dorfman & Glanzer, 1988; Mulligan, 2001; Zechmeister, Curt, & Sebastian, 1978; cf. DeCarlo, 2007). Whereas low-frequency words enjoy a benefit for recognition, high-frequency words tend to be remembered better when participants are required to generate a response from a set of items in memory (i.e., free recall; Balota & Neely, 1980; DeLosh & McDaniel, 1996; Gregg, 1976; Hall, 1954; Watkins, LeCompte, & Kim, 2000).

Interestingly, associative tasks like cued recall (Clark & Burchett, 1994; Criss, Aue, & Smith, 2011; Madan, Glaholt, & Caplan, 2010) and associative recognition (Chalmers & Humphreys, 2003; Clark, 1992; Clark & Burchett, 1994; Clark, Hori, & Callan, 1993; Clark & Shiffrin, 1992; Humphreys, et al., 2010) have also tended to demonstrate a high-frequency advantage, although the data are somewhat mixed (e.g., Hockley, 1994; Hockley & Cristi, 1996).² During associative recognition, participants are asked to discriminate *pairs* of items that were studied together (i.e., intact) from pairs containing items that were both studied, but in different pairings (i.e., rearranged). For example, if the pairs A–B and C–D are studied, an intact test pair would be A–B and a rearranged test pair would be A–D. To be successful at associative recognition, participants cannot simply rely on the familiarity of the individual items to make the decision, since both items were studied. Participants must recall that the items occurred *together* in a pair. In associative recognition, the HR to intact high-frequency pairs is usually equivalent to the HR for intact low-frequency pairs, but the FAR to rearranged low-frequency pairs is higher, resulting in an advantage for discriminating high-frequency pairs. The word frequency pattern reverses if the task includes pairs but does not require memory for associations. For pair recognition, in which A–B is discriminated from a pair of nonstudied items X–Y, the low-frequency advantage persists (Clark, 1992; Clark & Shiffrin, 1992).

¹ Recent evidence has suggested that this pattern may be more complicated than was once thought, especially when frequency is treated as a continuous rather than a dichotomous variable and when the frequency boundaries are pushed to their edges (Chalmers & Humphreys, 1998; Chalmers et al., 1997; Hemmer & Criss, 2013; Lohnas & Kahana, 2013).

² It is worth noting that there are differences between the cited studies in what constituted the low-frequency items, which could partially explain the mixed results.

There is disagreement about the mechanisms underlying the word frequency benefits in associative recognition. One possibility is that associative recognition may rely on recall processes (e.g., Dennis & Humphreys, 2001; Gronlund & Ratcliff, 1989; Rotello & Heit, 2000; Yonelinas, 1997), giving rise to the high-frequency benefit. Another possibility is that high-frequency items are easier to associate with other items (Criss et al., 2011; Gillund & Shiffrin, 1984; Madan et al., 2010), resulting in stronger associations between items.

Context variability

Given the surface similarity of word frequency and context variability, Steyvers and Malmberg (2003) investigated whether word frequency and context variability were functionally the same variable. To do so, they operationally defined context variability as the number of different documents in which a word appeared in the Touchstone Applied Science Associates (TASA) corpus (Landauer, Foltz, & Laham, 1998).³ From this, they extracted a subset of words that orthogonally combined high and low word frequency and variability. In an item recognition experiment, Steyvers and Malmberg observed mirror effects for both frequency *and* variability, in which items low in a property had a higher HR and a lower FAR than did those high in that property, a pattern that has been replicated elsewhere (e.g., Cook, Marsh, & Hicks, 2006). That is, word frequency effects were observed when context variability was held constant, and context variability effects were observed when word frequency was held constant. Steyvers and Malmberg concluded that word frequency and context variability represent different constructs that each contribute to memory. This conclusion has been validated across a number of different tasks. Low-context-variability advantages have been observed for source memory (Marsh, Cook, & Hicks, 2006a), lexical decision (e.g., Adelman, Brown, & Quesada, 2006; Hoffman & Woollams, 2015), word learning (Johns, Dye, & Jones, 2016), and serial recall (Parmentier, Comesaña, & Soares, 2017).

Particularly interesting is the observation that word frequency and context variability dissociate when the final test is free recall. High-frequency words tend to be recalled better than low-frequency words (e.g., Balota & Neely, 1980; DeLosh & McDaniel, 1996; Gregg, 1976; Hall, 1954; Watkins, LeCompte, & Kim, 2000). Conversely, *low*-context-variability words are recalled better than *high*-context-variability words when recalled freely. This pattern persists for both pure and mixed-composition lists (Hicks, Marsh &

³ Others have operationalized context variability (see also *contextual diversity* or *semantic diversity*) in this manner (e.g., Adelman, Brown, & Quesada, 2006), whereas others (e.g., Hoffman, Ralph, & Rogers, 2013; Jones, Johns, & Recchia, 2012) have focused more on semantic diversity.

Cook, 2005; Meeks, Knight, Brewer, Cook, & Marsh, 2014), when other word properties such as concreteness are controlled (Marsh, Meeks, Hicks, Cook, & Clark-Foos, 2006b), and across study durations (Cook et al., 2006).

Hicks et al. (2005) and Marsh, Meeks, et al. (2006b) hypothesized that the low-variability advantage was the result of fewer item-to-context associations for low-variability words. Appearing in fewer contexts presumably makes it easier to associate items to the experimental context. The researchers rationalized that this strong item-to-context binding improves all tests of episodic memory, because all episodic tests, by definition, require memory for the relevant context. Hicks et al. (2005) and Marsh, Meeks, et al. (2006b) went on to further suggest that if one were to disrupt the formation of item-to-context associations, this should also disrupt the low-context-variability advantage.

Indeed, Marsh, Meeks, et al. (2006b) were able to eliminate the low-context-variability advantage when participants were tested in environmental contexts different from those in the study condition (e.g., Smith, 1979). They concluded that removing the environmental context cues undercut the low-context-variability advantage by preventing reinstatement of those cues. Marsh, Meeks, et al., (2006b) eliminated the low-context-variability advantage in free recall when participants were asked to generate interitem associations during encoding. They concluded that focusing on interitem associations during encoding comes at the expense of item-to-context associations.

However, we question whether focusing on interitem associations necessarily comes at the expense of item-to-context associations. Criss, Aue, and Smith (2011) investigated the influences of context variability and word frequency on cued recall of word pairs, a task that necessitates associative encoding to be successful. When word frequency was held constant and the context variability of the cue and target were manipulated, Criss et al. observed a low-context-variability advantage for cues, contrary to Marsh, Meeks, et al.'s (2006b) theory. Given these results, it would seem that focusing on interitem associations during encoding did not eliminate the influence of context variability for a subsequent recall task. Thus, the influence of interitem associations on context variability effects is unresolved.

The present experiments

The aim of the present article is to address two questions:

- 1) Does focusing on associations during encoding harm performance for low-context-variability words during item recognition?
- 2) How does context variability influence performance on associative recognition?

To address Question 1, we examined whether focusing on interitem associations during encoding would disrupt the influence of context variability in recognition tasks by manipulating the test that participants expected to receive (e.g., Balota & Neely, 1980; Hockley & Cristi, 1996). To do so, we varied the training that participants received prior to test. In the item recognition training condition, participants were trained to expect a test of single-item recognition following study of the word pairs. Likewise, in the associative training condition, participants were trained to expect an associative recognition test. We also controlled the test that participants actually received (test condition), either an item recognition test or an associative recognition test. Word frequency and context variability were orthogonally manipulated during the final test block (Fig. 1). If Marsh, Meeks, et al.'s (2006b) hypothesis that interitem associations *disrupt* the low-context-variability advantage is correct, then we should observe context variability effects when the training focused on item information, and no effects of context variability when the training focused on associations. Put another way, we would expect training condition to interact with context variability.

As for Question 2, to our knowledge, this is the first article to explore the effect of context variability on associative recognition performance. Whereas context variability and word frequency behave similarly with respect to item recognition (Cook et al., 2006; Steyvers & Malmberg, 2003), they dissociate for free recall (Hicks et al., 2005; Marsh, Meeks, et al., 2006b) and for cued recall (Criss et al., 2011). If context variability dissociates from word frequency with regard to associative recognition, reflecting a low-context-variability advantage, this would provide further credence to the idea that these are distinct properties. Regardless, it is important to understand the role context that variability plays in associative recognition in order to constrain theories of memory.

Experiment 1

Method

Participants The participants in the experiment were 182 undergraduates from the Syracuse University research participation pool. Participants received course credit or \$10 per hour for their participation. The Syracuse University Institutional Review Board approved all study protocols, and all participants provided written informed consent.

Materials The word pool used for the training condition study–test cycles contained 408 medium-frequency words from 3 to 13 letters in length ($M = 5.93$, $SD = 1.91$). The words ranged between log frequencies of 5.2 and 13.2 ($M = 8.88$, $SD = 1.13$) in the Hyperspace Analogue to Language corpus (HAL; Balota et al., 2007; Lund & Burgess, 1996). The word

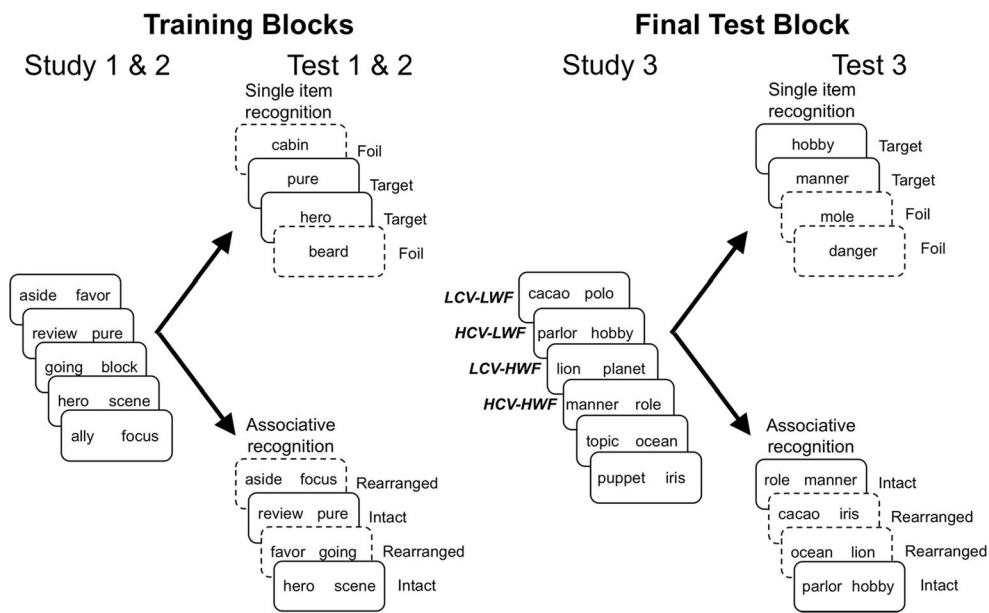


Fig. 1 Experimental design overview. Participants completed two study–test cycles with medium-frequency word pairs being tested for either item or associative recognition. They then received a third study–test cycle in which the word frequency (WF) and context variability (CV) of the word

pairs were manipulated, followed by a test of either item or associative recognition. The pairs were pure with respect to word frequency/context variability combination

pool used during the critical final study–test cycle to manipulate word frequency and context variability was the pool developed and used by Steyvers and Malmberg (2003), containing 288 nouns.⁴ The words ranged from 3 to 17 letters in length ($M = 7.11$, $SD = 2.21$ ⁵) and had been matched for orthographic distinctiveness across conditions by Steyvers and Malmberg.

Procedure Participants completed the experiment individually. The between-subjects manipulations were run as four separate experiments that crossed final test condition (associative or item recognition) with training condition (associative or item recognition). We combined the data for analysis under a single framework. During a session, participants received three study–test blocks in which they studied pairs of words. Unbeknownst to them, the first two blocks were included to train the participants to expect a particular test type. They were given two consecutive study–test blocks with either associative or item recogni-

tion as the test. They were not informed about the nature of the test in advance. Presumably, those given associative recognition would learn to focus on the associations between the studied items, and those given item recognition would focus on the individual items in order to maximize performance (Hockley & Cristi, 1996). The third block was the critical block, in which context variability and word frequency were manipulated within subjects. The study parameters were identical to those in the previous lists; however, the items were orthogonally manipulated to create pairs. Both items in the pair were randomly drawn from the same stimulus category (high or low frequency, high or low variability), and equal numbers of pairs from each category were studied.

During all study lists, participants studied 80 pairs of words, side by side for 3 s each, and were asked to rate the degree of association between the items on a 9-point scale (1 = *Not at all associated*, 9 = *Highly associated*). Testing was self-paced. During the item recognition test, participants were individually presented a total of 40 targets (randomly chosen from the studied items) and 40 foils, randomly intermixed. The foils were drawn randomly from the same corpus and in the same proportions as the targets. The task was to decide whether the presented item had been studied in the most recent list. During the associative recognition test, participants saw 40 intact (i.e., studied together) and 40 rearranged (i.e., studied separately) pairs. The task was to decide whether or not the items had been studied *together* in the previous list. For both the item and associative recognition tests, participants responded yes/no to make their recognition decisions. The entire experiment lasted a maximum of 60 min.

⁴ After data collection in Experiment 1 was complete, we discovered an error in the list construction, such that four words appeared among both the training and the context variability stimuli. All participants were impacted by at least one word appearing in both a training and a final study–test cycle. However, given the limited number of words affected and the large number of words that participants encountered in the randomized lists, there is no reason to believe that this error impacted the resulting performance. The error was corrected prior to Experiment 2.

⁵ Average word length for each frequency and variability condition: $M_{HV,HF} = 6.85$, $SD = 2.13$; $M_{HV,LF} = 8.17$, $SD = 2.35$; $M_{LV,HF} = 6.71$, $SD = 2.15$; $M_{LV,LF} = 6.74$, $SD = 1.89$

Design and analysis Training condition (item vs. associative expectation) and final test condition (item vs. associative recognition) were manipulated between subjects, whereas word frequency (high vs. low) and context variability (high vs. low) were manipulated within subjects. For clarity and ease of exposition, only analyses of discriminability (d') are discussed, but HRs and FARs are reported in their respective tables. Discriminability was calculated using $zFAR - zHR$ (Macmillan & Creelman, 1991). Edge correction was performed for calculating d' by replacing HR and FAR values of 0 with $0.5 \div n$ and values of 1 with $(n \div 5) \div n$, where n is the number of trials in the condition (Stanislaw & Todorov, 1999).

Analyses were performed in the R programming language (R Core Team, 2016) using the R packages ez (Lawrence, 2016), MBESS (Kelley, 2016), papaja (Aust & Barth, 2016), and BayesFactor (Morey & Rouder, 2015). Bayes factors and Cohen's d are provided for paired comparisons. Bayes factors represent the magnitude of the difference between the models under comparison. Bayes factors larger than 1 indicate greater evidence that the data arise from a model with a nonzero effect, whereas Bayes factors less than 1 indicate greater evidence for a null effect (i.e., zero difference) between the means.

Results

Replicating Steyvers and Malmberg (2003) First, we examined the data for the item recognition training condition and the item recognition final test condition. This was done simply to determine whether we had replicated the context variability effects observed previously (e.g., Steyvers & Malmberg, 2003).

As can be seen in Table 1, there was no interaction of word frequency and context variability [$F(1, 54) = 3.79, p$

$= .057, \eta_p^2 = .07$]; however, we observed main effects of both context variability [$F(1, 54) = 9.32, p = .004, \eta_p^2 = .15$] and word frequency [$F(1, 54) = 6.32, p = .015, \eta_p^2 = .10$]. Low-context-variability words were more discriminable than high-variability words ($d = 0.466, 95\% \text{ CI } [0.197, 0.733], BF_{10} = 26.69$), consistent with Steyvers and Malmberg (2003). Low-frequency words were more discriminable than high-frequency words ($d = 0.377, 95\% \text{ CI } [0.11, 0.643], BF_{10} = 4.2$). Next we evaluate the influences of encoding orientation and context variability on memory.

Context variability effects across training and test conditions

We performed a 2 (item vs. associative training) \times 2 (item vs. associative final recognition test) \times 2 (high vs. low frequency) \times 2 (high vs. low variability) mixed analysis of variance (ANOVA) for discriminability (d'). The data for HRs, FARs, and d' are presented in Tables 1 and 2.

For d' , we observed a three-way interaction of test condition, word frequency, and context variability [$F(1, 178) = 6.46, p = .012, \eta_p^2 = .04$; Figs. 2A and 2B]. An interaction of test condition and word frequency [$F(1, 178) = 26.38, p < .001, \eta_p^2 = .13$] is subsumed by this three-way interaction. The three-way interaction was driven by the fact that the interaction of word frequency and context variability switched as a function of test type. This is confirmed by examining the word frequency and context variability interactions separately for item recognition and associative recognition. For item recognition, we observed main effects of frequency [$F(1, 94) = 11.72, p = .001, \eta_p^2 = .11$] and variability [$F(1, 94) = 17.54, p < .001, \eta_p^2 = .16$], such that items low on each property enjoyed

Table 1 Experiment 1: Item recognition final test

	Item Recog. Training ($n = 55$)				Assoc. Recog. Training ($n = 40$)			
	High WF		Low WF		High WF		Low WF	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
Hit Rates								
High CV	.658	.033	.651	.026	.762	.026	.778	.027
Low CV	.695	.026	.665	.032	.800	.023	.770	.030
False Alarm Rates								
High CV	.302	.031	.253	.026	.238	.029	.198	.021
Low CV	.269	.032	.140	.026	.252	.031	.118	.017
d'								
High CV	1.33	0.154	1.47	0.170	1.83	0.152	2.03	0.164
Low CV	1.57	0.186	2.15	0.202	2.04	0.205	2.62	0.200

Word frequency (WF) and context variability (CV) effects in item recognition are shown under item or associative training with an associative encoding task

Table 2 Experiment 1: Associative recognition final test

	Item Recog. Training ($n = 38$)				Assoc. Recog. Training ($n = 49$)			
	High WF		Low WF		High WF		Low WF	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
Hit Rates								
High CV	.661	.031	.597	.032	.704	.033	.659	.031
Low CV	.716	.036	.634	.038	.729	.032	.663	.036
False Alarm Rates								
High CV	.189	.026	.239	.025	.206	.025	.182	.026
Low CV	.161	.029	.226	.029	.165	.026	.192	.023
d'								
High CV	1.70	0.212	1.21	0.166	1.93	0.215	1.84	0.186
Low CV	2.32	0.269	1.57	0.253	2.17	0.222	1.76	0.178

Word frequency (WF) and context variability (CV) effects in associative recognition are shown under item or associative training with an associative encoding task

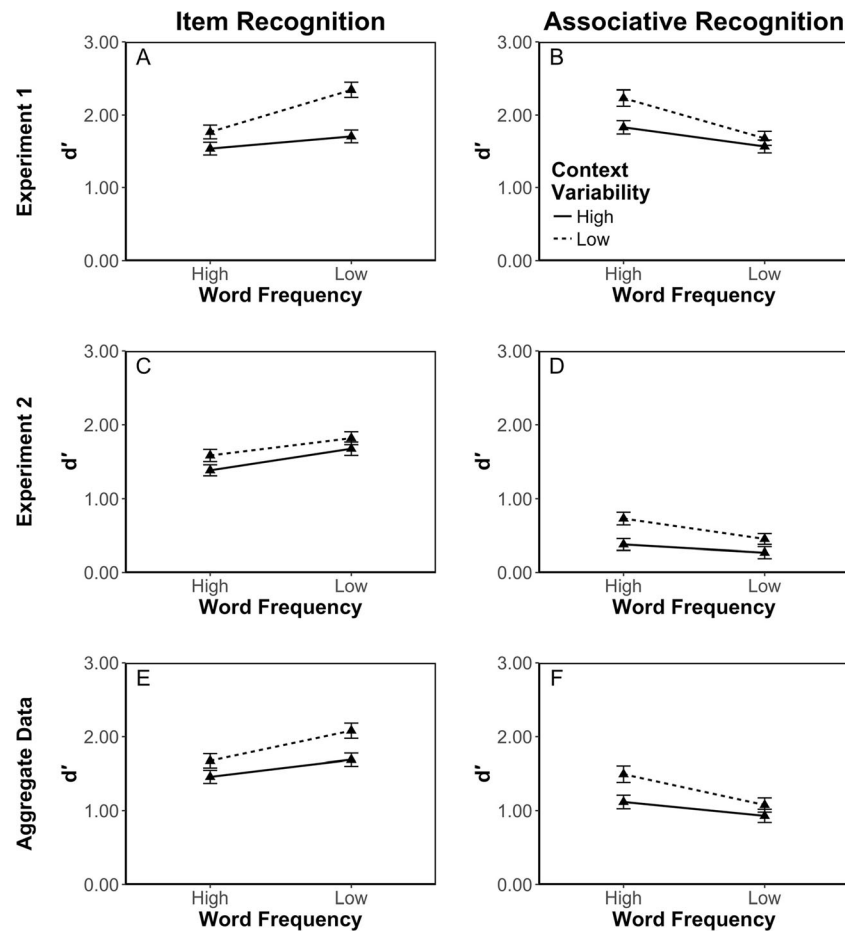


Fig. 2 Discriminability (d') data for each experiment are plotted as a function of word frequency (x-axis), context variability (dashed vs. solid lines), and final test condition (item vs. associative recognition; left and right panels, respectively). Low-context-variability words were consistently remembered better than high-context-variability words.

Word frequency and test condition interacted such that high-frequency words were remembered better in associative recognition, but low-frequency words were remembered better in item recognition. Error bars indicate within-subjects standard errors

an advantage, and an interaction of frequency and variability [$F(1, 94) = 5.28, p = .024, \eta_p^2 = .05$], in which performance was best for the combination of low frequency and variability. For associative recognition, we observed advantages for high-frequency [$F(1, 86) = 13.50, p < .001, \eta_p^2 = .14$] and low-variability [$F(1, 86) = 6.18, p = .015, \eta_p^2 = .07$] items, but no interaction [$F(1, 86) = 2.07, p = .154, \eta_p^2 = .02$]. This informs Question 2 that we posed earlier (*How does context variability influence performance on associative recognition?*). Low variability helps associative recognition, consistent with its influence on performance for other tasks. Furthermore, context variability and word frequency dissociate in their influences on associative recognition, providing converging evidence that the two properties are not functionally redundant.

We also observed main effects of training condition and context variability. For training condition, d' was higher under associative training [$F(1, 178) = 5.56, p = .019, d =$

$0.265, 95\% \text{ CI } [0.119, 0.411], \text{BF}_{10} = 40.14$]. For context variability, d' was higher for low-variability words [$F(1, 178) = 23.28, p < .001, d = 0.367, 95\% \text{ CI } [0.22, 0.513], \text{BF}_{10} = 7,429.93$]. The main effect of context variability and the fact that context variability did not interact with training condition inform Question 1 that we posed earlier (*Does focusing on associations during encoding harm performance for low-context-variability words during item recognition?*). According to Marsh, Meeks, et al. (2006b), we should have observed an interaction between context variability and training condition, such that low-context-variability words would benefit when trained on item but not on associative encoding. In contrast, we observed a low-context-variability advantage across the board, indicating that focusing on associations does not harm encoding for low-context-variability words (Fig. 3). No other effects on d' were significant ($F_s \leq 2.40, p_s \geq .123, \text{ and } \eta_p^2_s \leq .01$).

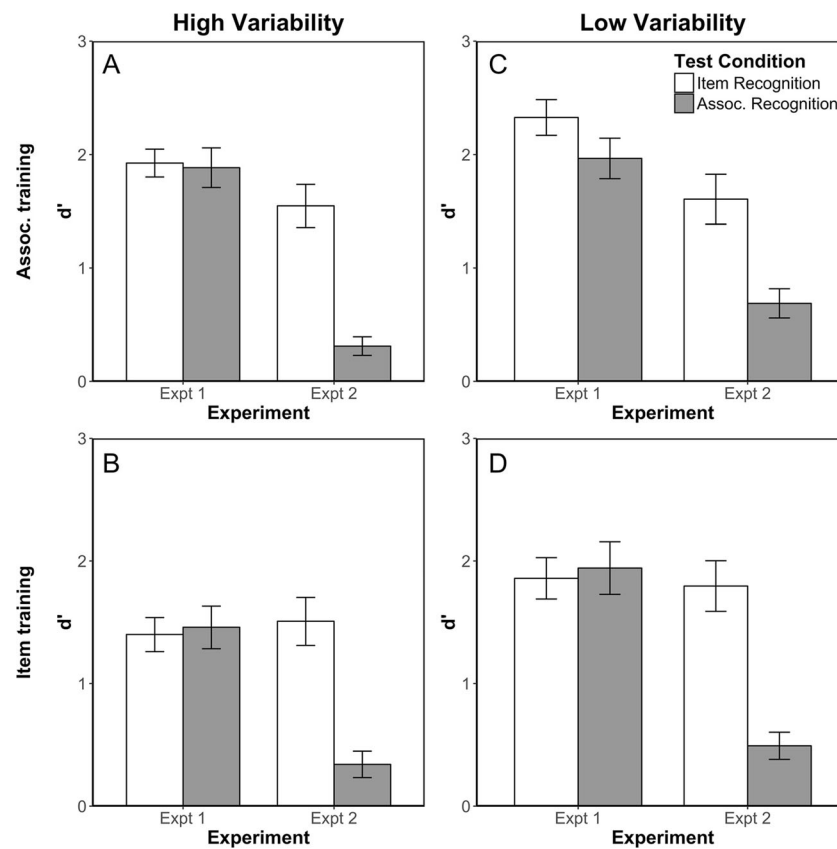


Fig. 3 Associative encoding (Exp. 1) resulted in better performance than item-based encoding (Exp. 2), and more so for associative recognition than for item recognition. Panels A, B, and D show this statistical interaction of experiment and final test condition. Panel C (low-variability

words in the associative training condition) is slightly different, showing no extra deficit for associative recognition testing, but similar degrees of harm for item and recognition testing

Discussion

In addressing the question of whether focusing on interitem associations during encoding comes at the expense of item-to-context binding (Question 1), we hypothesized that if focusing on associations attenuates the low-variability benefit, then we should observe a low-variability advantage under item recognition training, but a null context variability effect under associative recognition training. Instead, we observed an overall advantage for low-context-variability words. Anticipating an associative recognition test improved d' , but context variability did not interact with training condition. Thus, we did not find support in Experiment 1 for the idea that focusing on interitem associations comes at the expense of item-to-context associations.

When assessing context variability effects in associative recognition (Question 2), we observed evidence of a low-context-variability advantage in associative recognition. This is consistent with the low-variability advantage that has been observed across multiple tasks, including free recall (e.g., Marsh, Meeks, et al., 2006b). In contrast, we observed a high-frequency advantage in associative recognition, similar to that observed during free recall (e.g., Balota & Neely, 1980;

DeLosh & McDaniel, 1996; Gregg, 1976; Hall, 1954), which switched to a low-frequency advantage for the item recognition final test (e.g., Clark, 1992). The dissociation of word frequency and context variability for associative recognition provides credence to the idea that they are separate properties.

There are multiple differences between the present experiment and that of Marsh, Meeks, et al. (2006b). One possible criticism of Experiment 1 is that the encoding task did not allow for participants to efficiently process item information. In Experiment 1, all participants engaged in an associative orienting task by making an associative rating judgment for word pairs. The result could have been a disadvantage for the item encoding training condition. Perhaps the associative training group benefited from the congruency between the orienting task and training condition, leading to artificially inflated performance. The inflated performance could have masked the effects of training condition on context variability. In other words, we manipulated expectations by implicitly training participants on single-item or associative recognition. However, the encoding task present in all conditions encouraged associative encoding. We addressed this in Experiment 2 by having an encoding task that focused on item encoding.

Experiment 2

The aims and design for Experiment 2 were nearly identical to those of Experiment 1. The only substantive change was a change to the orienting task during encoding, to encourage participants to focus on processing the individual items within pairs. As with Experiment 1, if focusing on associations harms low-context-variability words selectively, via impaired item-to-context associations, then we would expect to find an interaction of training and context variability.

Method

Participants A total of 199 participants were recruited online through the Purdue University research participation pool. The participants received course credit for their participation, and the data were collected online. Twenty of the participants started but did not complete the experiment. Their data were dropped from the subsequent analysis, leaving 179 participants.

Materials The materials were as described in Experiment 1.

Procedure Experiment 2 was conducted online using the jsPsych (de Leeuw, 2015) JavaScript library. During encoding, participants studied pairs of words on each list, as in Experiment 1. They were asked to rate the pleasantness of one of the words in the pair. They did not know which word needed to be rated until after the pair had been removed from the screen. After the pair of words had disappeared, participants were asked to rate the pleasantness of either the left or the right word on a 9-point scale (1 = *Extremely Unpleasant*, 9 = *Extremely Pleasant*). Whether the left or the right word was prompted was randomized and evenly divided between left/right sides. All other aspects of Experiment 2 were identical to those of Experiment 1.

Results

Replicating Steyvers and Malmberg (2003) As with Experiment 1, we first examined the data for the item recognition training condition and the item recognition final test condition.

For d' , there was no interaction of word frequency and context variability [$F(1, 46) = 0.139, p = .711, \eta_p^2 < .01$], nor any effect of word frequency [$F(1, 46) = 1.91, p = .173, \eta_p^2 = .04$]. However, we observed a main effect of context variability [$F(1, 46) = 4.98, p = .031, \eta_p^2 = .10$], in which low-variability words were more discriminable than high-variability words ($d = 0.337, 95\% \text{ CI } [0.049, 0.625], \text{BF}_{10} = 1.43$), consistent with the previous results.

Context variability effects across training and test conditions

As before, we performed a 2 (item vs. associative training) \times 2 (item vs. associative final recognition test) \times 2 (high vs. low frequency) \times 2 (high vs. low variability) mixed ANOVA on discriminability (d'). The data for d' , HRs, and FARs are presented in Tables 3 and 4.

As can be seen in Figs. 2C and 2D, we observed an interaction of test condition and word frequency [$F(1, 175) = 11.07, p < .001, \eta_p^2 = .06$], reflecting the same word frequency crossover pattern observed in Experiment 1. That is, low-frequency words were remembered better for item recognition, but high-frequency pairs were remembered better in associative recognition.

We also observed main effects of context variability [$F(1, 175) = 12.30, p < .001, \eta_p^2 = .07$] and test condition [$F(1, 175) = 54.19, p < .001, \eta_p^2 = .24$]. Low-context-variability words were recognized better than high-variability words, as in Experiment 1 ($d = 0.268, 95\% \text{ CI } [0.121, 0.415], \text{BF}_{10} = 31.10$). Overall, we replicated Experiment 1 in terms of the roles of word frequency and context variability in test condition: Word frequency effects were moderated by final test condition, whereas context variability effects are not.

Performance on associative recognition using an item-encoding task was abysmal (consistent with, e.g., Hockley & Cristi, 1996; Murdock, 1982). This is evidenced by the main effect of test condition, reflecting substantially better performance for item recognition than for associative recognition ($d = 0.929, 95\% \text{ CI } [0.774, 1.08], \text{BF}_{10} = 1.69 \times 10^{27}$). No other effects or interactions for d' were significant ($F_s \leq 3.30, p_s \geq .071$, and $\eta_p^2_s \leq .02$).

Table 3 Experiment 2: Item recognition final test

	Item Recog. Training ($n = 47$)		Assoc. Recog. Training ($n = 48$)					
	High WF	Low WF	High WF	Low WF	High WF	Low WF		
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
Hit Rates								
High CV	.698	.032	.687	.032	.717	.030	.765	.029
Low CV	.753	.027	.732	.030	.721	.035	.729	.035
False Alarm Rates								
High CV	.345	.037	.260	.032	.348	.030	.317	.031
Low CV	.287	.029	.243	.035	.348	.031	.273	.033
d'								
High CV	1.38	0.215	1.64	0.219	1.38	0.201	1.71	0.215
Low CV	1.71	0.201	1.88	0.255	1.46	0.236	1.76	0.237

Word frequency (WF) and context variability (CV) effects in item recognition are shown under item or associative training with an item-based encoding task

Table 4 Experiment 2: Associative recognition final test

	Item Recog. Training (<i>n</i> = 41)				Assoc. Recog. Training (<i>n</i> = 43)			
	High WF		Low WF		High WF		Low WF	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
Hit Rates								
High CV	.478	.028	.554	.032	.479	.033	.477	.033
Low CV	.576	.032	.598	.029	.577	.029	.579	.031
False Alarm Rates								
High CV	.390	.028	.437	.032	.358	.036	.405	.028
Low CV	.368	.030	.505	.030	.365	.034	.426	.032
<i>d'</i>								
High CV	0.323	0.128	0.335	0.141	0.433	0.141	0.185	0.106
Low CV	0.664	0.143	0.319	0.133	0.793	0.168	0.580	0.147

Word frequency (WF) and context variability (CV) effects in associative recognition are shown under item or associative training with an item-based encoding task

Discussion

The purpose of this experiment was to replicate Experiment 1 using an encoding task that encouraged item encoding. The general patterns of data from Experiment 1 were replicated. With regard to Question 2 posed earlier, accuracy was higher for low-context-variability words than for high variability in both item and associative recognition, replicating Experiment 1. Moreover, training participants to engage in item or associative recognition did not change the low-context-variability advantage, addressing Question 1 posed earlier. This further substantiates the idea that context variability and word frequency are related but separate lexical properties. These data add to the growing literature showing that the low-context-variability advantage persists across tasks, unlike word frequency.

Associative recognition performance was very poor, as one might expect given the encoding task (e.g., Hockley & Cristi, 1996; Murdock, 1982). Although both experiments had robust sample sizes, it could be useful to combine the data from the experiments to look for emergent patterns across the data and to further evaluate the role of encoding task separately from training condition.

Comparison of Experiments 1 and 2

Experiments 1 and 2 employed different encoding tasks. In Experiment 1, participants made associative judgments about the pairs during encoding, whereas in Experiment 2 they made pleasantness judgments about a single item in each pair. Thus far, we have examined performance within experiments with regard to our questions of interest. It may be informative to

also aggregate the experimental data in order to examine emergent patterns across the experiments and compare the encoding tasks directly.

To examine the impact of encoding task, we aggregated the data for Experiments 1 and 2 and coded for experimental group. We performed a 2 (item vs. associative training) × 2 (item vs. associative final recognition test) × 2 (high vs. low frequency) × 2 (high vs. low variability) × 2 (Exp. 1 vs. Exp. 2) mixed ANOVA on *d'*.

We observed main effects of experiment [$F(1, 353) = 54.46, p < .001, \eta_p^2 = .13$], final test condition [$F(1, 353) = 31.00, p < .001, \eta_p^2 = .08$], and context variability [$F(1, 353) = 35.34, p < .001, \eta_p^2 = .09$]. For the main effect of experiment, the item-focused orientation clearly harmed performance relative to the associative orienting task in Experiment 1. Overall, *d'* was better for Experiment 1 (associative encoding) than for Experiment 2 (item encoding; $d = 0.55, 95\% \text{ CI } [0.445, 0.655], \text{BF}_{10} = 2.91 \times 10^{21}$). For the main effect of final test condition, *d'* was higher for item recognition than for associative recognition ($d = 0.407, 95\% \text{ CI } [0.302, 0.511], \text{BF}_{10} = 1.99 \times 10^{11}$). For the main effect of context variability, low-variability items were more discriminable than high-variability items ($d = 0.32, 95\% \text{ CI } [0.217, 0.424], \text{BF}_{10} = 2,753,140$).

We again observed the three-way interaction of final test condition, word frequency, and context variability [$F(1, 353) = 4.99, p = .026, \eta_p^2 = .01$; Figs. 2E and 2F], in which low-context-variability words were remembered better across the board, but word frequency effects depended on the test condition. Final test condition and word frequency interacted [$F(1, 353) = 36.67, p < .001, \eta_p^2 = .09$], reflecting the word frequency crossover effect, in which high frequency was superior in associative recognition but low frequency was superior in item recognition. A final test condition and experiment interaction [$F(1, 353) = 24.75, p < .001, \eta_p^2 = .07$] reflected the exceptionally poor performance for associative recognition under emphasis on encoding individual items.

Together, these effects form a four-way interaction between experiment, training condition, final test condition, and context variability [$F(1, 353) = 4.24, p = .04, \eta_p^2 = .01$] that was significant by traditional frequentist measures. We interpret this interaction at the request of the reviewers; however, we note that the effect is small, the *p* value is uncertain for a post-hoc analysis, and interpreting such interactions is not without concern (e.g., Loftus, 1978; Wagenmakers, Kryptos, Criss, & Iverson, 2012). Figure 3 helps dissect the four-way interaction. The effect appears to be driven by the fact that a three-way interaction is present for the associative recognition training condition that is absent for item recognition training. For associative recognition training, a three-way interaction of experiment, final test condition, and context variability is evident. For high-variability words, there is clearly an interaction

of experiment and final test condition, reflecting poorer performance on associative recognition tests in Experiment 2 (Fig. 3A). This same pattern is also evident in the item recognition training conditions in both experiments (Fig. 3B, 3D). For low-variability words in the associative recognition training condition, there are interpretable main effects of final test condition (item recognition > associative recognition) and experiment (Exp. 1 > Exp. 2; Fig. 3C). Low-variability words in Experiment 1 seem to have benefited from the combination of encoding condition and associative training for item recognition in particular. If it were the case that focusing on associations disrupted the low-variability advantage, this should have been evident when participants had both associative encoding and associative training tasks. No other effects on d' were significant ($F_s \leq 2.85$, $p_s \geq .092$, and $r_p^2 \leq .01$).

To summarize, we observed a persistent low-context-variability advantage across experiments. If focusing on interitem associations harms performance for low-variability items, we did not find evidence of it. Context variability interacted with training condition as part of a four-way interaction; however, the context variability effect was not attenuated. In fact, low-variability items appeared to benefit from the combination of the associative encoding task and the associative training condition, particularly for item recognition. Put simply, the associative encoding task did not attenuate the context variability benefit. This is different from the word frequency effect, which traded off as a function of final test.

General discussion

The aim of the present set of experiments was twofold, as was described by the questions posed in the introduction. With Question 1, we asked whether focusing on associations during encoding attenuated the low-context-variability advantage, as had been observed by Marsh, Meeks, et al. (2006b). To address this, within each experiment, participants were trained to focus on either item or associative testing. Across experiments, we manipulated whether participants employed an associative or an item orienting task during encoding. With Question 2, we asked whether low-context-variability effects persisted for associative recognition.

Marsh, Meeks, et al. (2006b) predictions

We first sought to test the prediction that focusing on associations attenuates the low-context-variability advantage. This would have presented as an interaction of training condition and context variability in which the low-variability advantage would be attenuated under associative training. Alternatively, we might have observed a low-context-variability advantage when encoding was focused on items (Exp. 2) but not on

associations (Exp. 1). In both experiments, performance was better for low-context-variability words than for high-context-variability words, and focusing on associations did not change that result.

Marsh, Meeks, et al. (2006b) observed a null context variability effect in free recall when participants engaged in an associative encoding task for individual items. In their experiment, individual items were presented sequentially, and participants were asked to associate the current item with the immediately preceding item by reporting a similarity between the words. Marsh, Meeks, et al. suggested that focusing on interitem associations during encoding weakened the item-to-context associations, attenuating the low-context-variability advantage (e.g., Hicks et al., 2005).

Given the present results and those of Criss et al. (2011) using cued recall, it is apparent that focusing on associations does not attenuate the low-context-variability advantage. In the present data, we also observed that focusing on associations or employing an associative encoding task improved performance. We suggest that high-context-variability words may have benefited from the associative encoding task in Marsh, Meeks, et al.'s (2006b) data, resulting in a null context variability effect. Consider the data presented in Fig. 4. In the figure, we aggregated performance for low- and high-variability words in the control conditions of Marsh, Meeks, et al.'s Experiment 2 (their Fig. 2) and Experiment 3 (their Fig. 3). Marsh, Meeks, et al. manipulated the environmental context using either music (Exp. 2) or physical location (Exp. 3); however, here we focus on the control conditions, in which encoding/retrieval context was constant. Notably, the free recall performance in the control (no music) condition for Experiment 2 was worse for high-context-variability words ($M \approx .248$) than for low-context-variability words ($M \approx .328$), a typical context variability effect. However, when an associative encoding task was employed (Exp. 3), a null context variability effect is observed; correct performance for low-context-variability ($M \approx .329$) words was similar to that for high-context-variability words ($M \approx .328$). In Marsh, Meeks, et al.'s Experiment 2, the null context variability effect was not the product of attenuated low-variability performance, but appears to be the result of improved performance for high-context-variability words.

Test expectancy and encoding tasks

It is not altogether unexpected that an associative encoding task should improve memory performance. Indeed, many tasks, such as considering semantic relevance (Craik & Lockhart, 1972; Craik & Tulving, 1975), forming interactive images (Bower, 1970), evaluating self relevance (Rogers, Kuiper, & Kirker, 1977), or rating items for survival relevance (Nairne, Thompson, & Pandeirada, 2007), amongst others, have been reported to improve memory for the studied

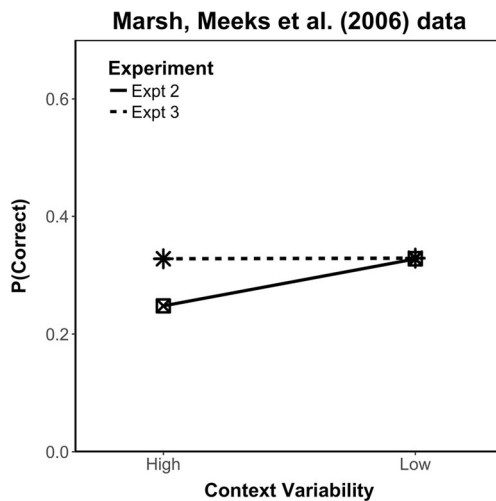


Fig. 4 Data adapted from Marsh, Meeks, et al. (2006b) depicting performance on a free recall test for the control conditions of their Experiments 2 and 3. Experiment 3 employed an associative encoding task, and there they observed a null context variability effect (asterisks), whereas a low-context-variability advantage was observed in Experiment 2 (boxes) when the associated encoding task was not used. Their observed null context variability effect appears to result from an improvement in high-context-variability recall during associative encoding. The relevant data were extracted from the Marsh, Meeks, et al. figures using Data Thief 3 (version 1.6)

materials. In the present experiments, the associative encoding task in Experiment 1 was associated with substantially better performance than was the item encoding task in Experiment 2, particular for the associative recognition test.

When left to their own devices, participants can also alter their own encoding strategies, depending on the anticipated demands of a memory test (Finley & Benjamin, 2012). Experiments examining such test expectancy effects typically contrast participants anticipating a test in which they must generate the items that were studied (i.e., free recall) or decide whether or not a presented item had been studied (i.e., recognition). The results of the influence of test expectancy are mixed. In some cases, participants perform best on the test that they expected (Tversky, 1973), perhaps suggesting task-specific processes. However, a somewhat more common observation is that performance tends to be better for both tasks when participants anticipate a recall test relative to a recognition test (Balota & Neely, 1980; Neely & Balota, 1981). Interestingly, anticipating a recall test appears to improve performance for high-frequency items in particular (Balota & Neely, 1980), similar to the way that the associative encoding task appears to have improved performance for high-variability items in Marsh, Meeks, et al.'s (2006b).

In the present experiments, we observed that engaging in an associative encoding (Exp. 1) was associated with improved performance overall. However, engaging in the associative encoding task was not associated with an attenuation of the low-context-variability advantage. Likewise, manipulating test expectancy via the associative recognition training condition

improved performance; however, the training benefit was smaller than the encoding task manipulation and was specific to Experiment 1. The weakness of the test expectancy (training) manipulation that we employed could be attributed to differences from other test expectancy experiments. We used fewer study–test training cycles to induce test expectancy than have other experiments. For example, both Finley and Benjamin (2012) and Cho and Neely (2017) used four study–test training cycles to induce test expectancy. Additionally, instead of a recall test, we contrasted anticipation of an associative recognition test with an item recognition test. Although it could be the case that associative recognition relies on aspects of recall processes (e.g., Clark, 1992; Clark & Shiffrin, 1992), it seems reasonable to suggest that participants could approach preparing for a free recall task differently from preparing for an associative recognition task.

For the associative encoding benefits that we observed, we can speculate about the potential benefits of focusing on associative information during encoding. Perhaps the associative orienting task improved performance by making it easier to bind to the experimental context. For example, participants may have engaged in some form of interactive imagery of the to-be-remembered words during encoding. Such imagery has been demonstrated to improve memorability (Bower, 1970). Perhaps such an image would effectively reduce context noise for high-frequency items by making a unique image that is easier to bind to the experimental context. Put another way, it could be that focusing on associations selectively improves the encoding of associative features, thereby strengthening item-to-context associations. Another possibility is simply that focusing on associations or employing an associative encoding task led to more complete encoding for the pair of items. The more complete representation thus afforded access to both item and associative information.

Interplay of lexical properties

The present data provide another instance in which word frequency and context variability dissociate, namely for associative recognition. When the final test was item recognition, low-word-frequency, low-context-variability words enjoyed the greatest benefit. However, when the test was associative recognition, the word frequency advantage switched to high-frequency words, whereas the low-context-variability advantage persisted; high-frequency, low-context-variability words were the most discriminable. This is consistent with the word frequency crossover effect observed elsewhere (Clark, 1992). Thus, the combination of word frequency and context variability is what determined performance, and these variables appear to do so in reliably different ways. However, if context variability and word frequency were simply the same construct, as

some models of memory might suggest (e.g., Buchler & Reder, 2007; Dennis & Humphreys, 2001; Reder et al., 2000), then one would not expect them to dissociate across tasks.

We suggest that perhaps word frequency and context variability share separate, complementary roles in memory. As we mentioned earlier, explanations of word frequency effects are plentiful and vary depending on the model in question. The “retrieving effectively from memory” model (REM; Shiffrin & Steyvers, 1997) explains the low-frequency benefit in item recognition as those words having different features (e.g., Diana & Reder, 2006; McClelland & Chappell, 1998) that are highly diagnostic (e.g., Criss & Malmberg, 2008), whereas the search of associative memory model (Gillund & Shiffrin, 1984; Raaijmakers & Shiffrin, 1981) explains better memory for high-frequency words because they are easier to associate with other words. However, the question then becomes, what unique role could context variability play in a memory search?

The notion of the strategic use of context information is common in models of recall (e.g., Howard & Kahana, 2002; Kılıç, Criss, & Howard, 2013; Lehman & Malmberg, 2009, 2013; Malmberg & Shiffrin, 2005; Polyn, Norman, & Kahana, 2009) but is less common for recognition (Dennis & Humphreys, 2001; cf. Malmberg, 2008). Recognition models in which context plays a central role (e.g., Dennis & Humphreys, 2001) could account for context variability effects, indicating that a low-context-variability advantage is derived from the lack of interference from other contexts. However, the issue is complicated by the conflation of word frequency and context variability properties in such models. Representing item and context information separately in a model, as is the case in REM (Shiffrin & Steyvers, 1997), might be able to handle the constellation of observed word frequency and context variability data. In REM, context is represented by a set of features similar to item and semantic features.

It is conceivable that, similar to word frequency, items associated with relatively fewer contexts tend to have more distinctive context features. In this way, context could be used to select the appropriate episode to which the test item is compared (in recognition tasks) or from which the item is retrieved. By virtue of the many contexts in which high-context-variability words have been experienced, this process is less accurate for high-context-variability words, leading to a less-constrained activated set and lower performance for high-context-variability words. This context match process would benefit both recognition and recall tasks, which would explain the persistent low-context-variability benefit observed. However, this is conjecture, and the question will remain open until a concrete computational account is provided.

Summary

At the outset, we sought to answer two questions:

- 1) Does focusing on associations harm performance for low-context-variability words during item recognition?
- 2) How does context variability influence performance on associative recognition?

We found no evidence that focusing on interitem associations, either through encoding task or training for associative (vs. item) recognition, altered the low-context-variability advantage. There was a persistent low-context-variability advantage across both item and associative recognition. We also observed a dissociation in the influences of word frequency and context variability on associative recognition. When memory was tested using item recognition, a typical low-frequency advantage was observed, whereas a test of associative recognition demonstrated a high-frequency advantage. This reinforces the idea that word frequency and context variability are distinct properties of words that inform separate aspects of the memory search process. We suggest that word frequency differentially affects encoding, in that diagnostic item features are stored for low-frequency words and associative features are more likely to be stored for high-frequency words. Context variability plays a role during retrieval, with high-context-variability words, to their detriment, activating a more diverse, less-constrained collection of episodic memories that play a role in the eventual memory decision. Replicating Hockley and Cristi (1996), we found an advantage for both tasks when associative recognition was expected, suggesting that a focus on associative information improves encoding.

Author note This material is partially based upon work supported by the National Science Foundation under Grant No. 0951612, awarded to A.H.C. J.M.F. was partially supported by the Psychology Research Initiative in Diversity Enhancement program in the Department of Psychology at Syracuse University. The data and materials can be found at <https://osf.io/3y7xp>.

References

- Adelman, J. S., Brown, G. D. A., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science, 17*, 814–823.
- Aust, F., & Barth, M. (2016). Papaja: Create APA manuscripts with R Markdown. (v. 0.1.0.9492). Retrieved from <https://github.com/crsh/papaja>
- Balota, D. A., & Neely, J. H. (1980). Test-expectancy and word-frequency effects in recall and recognition. *Journal of*

- Experimental Psychology: Human Learning and Memory*, 6, 576–587. <https://doi.org/10.1037/0278-7393.6.5.576>
- Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., . . . Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, 39, 445–459. <https://doi.org/10.3758/BF03193014>
- Bower, G. H. (1970). Imagery as a relational organizer in associative learning. *Journal of Verbal Learning and Verbal Behavior*, 9, 529–533.
- Buchler, N. E., & Reder, L. M. (2007). Modeling age-related memory deficits: A two-parameter solution. *Psychology and Aging*, 22, 104–121.
- Chalmers, K. A., & Humphreys, M. S. (1998). Role of generalized and episode specific memories in the word frequency effect in recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 610–632.
- Chalmers, K., & Humphreys, M. (2003). Experimental manipulation of prior experience: Effects on item and associative recognition. *Memory*, 11, 233–246.
- Chalmers, K. A., Humphreys, M. S., & Dennis, S. (1997). A naturalistic study of the word frequency effect in episodic recognition. *Memory & Cognition*, 25, 780–784.
- Cho, K. W., & Neely, J. H. (2017). The roles of encoding strategies and retrieval practice in test-expectancy effects. *Memory*, 25, 626–635.
- Clark, S. E. (1992). Word frequency effects in associative and item recognition. *Memory & Cognition*, 20, 231–243.
- Clark, S. E., & Burchett, R. E. (1994). Word frequency and list composition effects in associative recognition and recall. *Memory & Cognition*, 22, 55–62.
- Clark, S. E., Hori, A., & Callan, D. E. (1993). Forced-choice associative recognition: Implications for global-memory models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 871–881.
- Clark, S. E., & Shiffrin, R. M. (1992). Cuing effects and associative information in recognition memory. *Memory & Cognition*, 20, 580–598.
- Cook, G. I., Marsh, R. L., & Hicks, J. L. (2006). The role of recollection and familiarity in the context variability mirror effect. *Memory & Cognition*, 34, 240–250. <https://doi.org/10.3758/BF03193402>
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of verbal learning and verbal behavior*, 11, 671–684. [https://doi.org/10.1016/S0022-5371\(72\)80001-X](https://doi.org/10.1016/S0022-5371(72)80001-X)
- Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, 104, 268–294. <https://doi.org/10.1037/0096-3445.104.3.268>
- Criss, A. H., Aue, W. R., & Smith, L. (2011). The effects of word frequency and context variability in cued recall. *Journal of Memory and Language*, 64, 119–132. <https://doi.org/10.1016/j.jml.2010.10.001>
- Criss, A. H., & Malmberg, K. J. (2008). Evidence in favor of the early-phase elevated-attention hypothesis: The effects of letter frequency and object frequency. *Journal of Memory and Language*, 59, 331–345.
- Criss, A. H., & Shiffrin, R. M. (2004). Interactions between study task, study time, and the low-frequency hit rate advantage in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 778–786. <https://doi.org/10.1037/0278-7393.30.4.778>
- DeCarlo, L. T. (2007). The mirror effect and mixture signal detection theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 18–33. <https://doi.org/10.1037/0278-7393.33.1.18>
- de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, 47, 1–12. <https://doi.org/10.3758/s13428-014-0458-y>
- DeLosh, E. L., & McDaniel, M. A. (1996). The role of order information in free recall: Application to the word-frequency effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 1136–1146. <https://doi.org/10.1037/0278-7393.22.5.1136>
- Dennis, S., & Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review*, 108, 452–478. <https://doi.org/10.1037/0033-295X.108.2.452>
- Diana, R. A., & Reder, L. M. (2006). The low-frequency encoding disadvantage: Word frequency affects processing demands. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32, 805–815. <https://doi.org/10.1037/0278-7393.32.4.805>
- Dorfman, D., & Glanzer, M. (1988). List composition effects in lexical decision and recognition memory. *Journal of Memory and Language*, 27, 633–648.
- Finley, J. R., & Benjamin, A. S. (2012). Adaptive and qualitative changes in encoding strategy with experience: evidence from the test-expectancy paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38, 632–652.
- Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, 91, 1–67. <https://doi.org/10.1037/0033-295X.91.1.1>
- Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition*, 13, 8–20. <https://doi.org/10.3758/BF03198438>
- Glanzer, M., & Adams, J. K. (1990). The mirror effect in recognition memory: Data and theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 5–16. <https://doi.org/10.1037/0278-7393.16.1.5>
- Glanzer, M., & Bowles, N. (1976). Analysis of the word-frequency effect in recognition memory. *Journal of Experimental Psychology: Human Learning and Memory*, 2, 21–31.
- Gregg, V. (1976). Word frequency, recognition and recall. In J. Brown (Ed.), *Recall and recognition* (pp. 183–216). Oxford, UK: Wiley.
- Gronlund, S. D., & Ratcliff, R. (1989). Time course of item and associative information: Implications for global memory models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 846–858.
- Hall, J. F. (1954). Learning as a function of word-frequency. *American Journal of Psychology*, 67, 138–140.
- Hemmer, P., & Criss, A. H. (2013). The shape of things to come: Evaluating word frequency as a continuous variable in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39, 1947–1952. <https://doi.org/10.1037/a0033744>
- Hicks, J. L., Marsh, R. L., & Cook, G. I. (2005). An observation on the role of context variability in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 1160–1164.
- Hintzman, D. L., Caulton, D. A., & Curran, T. (1994). Retrieval constraints and the mirror effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 275–289. <https://doi.org/10.1037/0278-7393.20.2.275>
- Hockley, W. E. (1994). Reflections of the mirror effect for item and associative recognition. *Memory & Cognition*, 22, 713–722.
- Hockley, W. E., & Cristi, C. (1996). Tests of encoding tradeoffs between item and associative information. *Memory & Cognition*, 24, 202–216. <https://doi.org/10.3758/BF03200881>
- Hoffman, P., Ralph, M. A. L., & Rogers, T. T. (2013). Semantic diversity: A measure of semantic ambiguity based on variability in the contextual usage of words. *Behavior Research Methods*, 45, 718–730.
- Hoffman, P., & Woollams, A. M. (2015). Opposing effects of semantic diversity in lexical and semantic relatedness decisions. *Journal of Experimental Psychology: Human Perception and Performance*, 41, 385–402. <https://doi.org/10.1037/a0038995>
- Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, 46, 269–299. <https://doi.org/10.1006/jmps.2001.1388>
- Humphreys, M. S., Maguire, A. M., McFarlane, K. A., Burt, J. S., Bolland, S. W., Murray, K. L., & Dunn, R. (2010). Using maintenance rehearsal to explore recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 147–159. <https://doi.org/10.1037/a0017687>

- Johns, B. T., Dye, M., & Jones, M. N. (2016). The influence of contextual diversity on word learning. *Psychonomic Bulletin & Review*, 23, 1214–1220.
- Jones, M. N., Johns, B. T., & Recchia, G. (2012). The role of semantic diversity in lexical organization. *Canadian Journal of Experimental Psychology*, 66, 115–124. <https://doi.org/10.1037/a0026727>
- Karlsen, P., & Snodgrass, J. (2004). The word-frequency paradox for recall/recognition occurs for pictures. *Psychological Research*, 68, 271–276.
- Kelley, K. (2016). MBESS: The MBESS R package (Version 4.1.0). Retrieved from <https://CRAN.R-project.org/package=MBESS>
- Kılıç, A., Criss, A. H., & Howard, M. W. (2013). A causal contiguity effect that persists across time scales. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39, 297–303. <https://doi.org/10.1037/a0028463>
- Landauer, T. K., Foltz, P. W., & Laham, D. (1998). Introduction to latent semantic analysis. *Discourse Processes*, 25, 259–284. <https://doi.org/10.1080/01638539809545028>
- Lawrence, M. A. (2016). ez: Easy analysis and visualization of factorial experiments (Version 4.4-0). Retrieved from <https://CRAN.R-project.org/package=ez>
- Lehman, M., & Malmberg, K. J. (2009). A global theory of remembering and forgetting from multiple lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35, 970–988. <https://doi.org/10.1037/a0015728>
- Lehman, M., & Malmberg, K. J. (2013). A buffer model of memory encoding and temporal correlations in retrieval. *Psychological Review*, 120(1), 155–189.
- Loftus, G. R. (1978). On interpretation of interactions. *Memory & Cognition*, 6, 312–319. <https://doi.org/10.3758/BF03197461>
- Lohnas, L. J., & Kahana, M. J. (2013). Parametric effects of word frequency in memory for mixed frequency lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39, 1943–1946.
- Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers*, 28, 203–208. <https://doi.org/10.3758/BF03204766>
- Macmillan, N., & Creelman, C. (1991). *Detection Theory: A User's Guide*. New York, NY: Cambridge University Press.
- Madan, C. R., Glaholt, M. G., & Caplan, J. B. (2010). The influence of item properties on association-memory. *Journal of Memory and Language*, 63, 46–63.
- Malmberg, K. J. (2008). Recognition memory: A review of the critical findings and an integrated theory for relating them. *Cognitive Psychology*, 57, 335–384.
- Malmberg, K. J., & Shiffrin, R. M. (2005). The “one-shot” hypothesis for context storage. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 322–336. <https://doi.org/10.1037/0278-7393.31.2.322>
- Marsh, R. L., Cook, G. I., & Hicks, J. L. (2006a). The effect of context variability on source memory. *Memory & Cognition*, 34, 1578–1586. <https://doi.org/10.3758/BF03195921>
- Marsh, R. L., Meeks, J. T., Hicks, J. L., Cook, G. I., & Clark-Foos, A. (2006b). Concreteness and item-to-list context associations in the free recall of items differing in context variability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32, 1424–1430.
- McClelland, J. L., & Chappell, M. (1998). Familiarity breeds differentiation: A subjective-likelihood approach to the effects of experience in recognition memory. *Psychological Review*, 105, 724–760. <https://doi.org/10.1037/0033-295X.105.4.734-760>
- Meeks, J. T., Knight, J. B., Brewer, G. A., Cook, G. I., & Marsh, R. L. (2014). Investigating the subjective reports of rejection processes in the word frequency mirror effect. *Consciousness and Cognition*, 24, 57–69.
- Morey, R. D., & Rouder, J. N. (2015). BayesFactor: Computation of Bayes factors for common designs (Version 0.9.12-2). Retrieved from <https://CRAN.R-project.org/package=BayesFactor>
- Mulligan, N. W. (2001). Word frequency and memory: Effects on absolute versus relative order memory and on item memory versus order memory. *Memory & Cognition*, 29, 977–985.
- Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, 89, 609–626. <https://doi.org/10.1037/0033-295X.89.6.609>
- Naime, J. S., Thompson, S. R., & Pandeirada, J. N. (2007). Adaptive memory: survival processing enhances retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 263–273. <https://doi.org/10.1037/0278-7393.33.2.263>
- Neely, J. H., & Balota, D. A. (1981). Test-expectancy and semantic-organization effects in recall and recognition. *Memory & Cognition*, 9, 283–300. <https://doi.org/10.3758/BF03196962>
- Parmentier, F. B., Comesaña, M., & Soares, A. P. (2017). Disentangling the effects of word frequency and contextual diversity on serial recall performance. *Quarterly Journal of Experimental Psychology*, 70, 1–17.
- Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review*, 116, 129–156. <https://doi.org/10.1037/a0014420>
- R Core Team. (2016). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from www.R-project.org/.
- Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, 88, 93–134. <https://doi.org/10.1037/0033-295X.88.2.93>
- Rao, K. V., & Proctor, R. W. (1984). Study-phase processing and the word frequency effect in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10, 386–394.
- Reder, L. M., Nhouyvanisvong, A., Schunn, C. D., Ayers, M. S., Angstadt, P., & Hiraki, K. (2000). A mechanistic account of the mirror effect for word frequency: A computational model of remember-know judgments in a continuous recognition paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 294–320. <https://doi.org/10.1037/0278-7393.26.2.294>
- Rogers, T. B., Kuiper, N. A., & Kirker, W. S. (1977). Self-reference and the encoding of personal information. *Journal of Personality and Social Psychology*, 35, 677–688. <https://doi.org/10.1037/0022-3514.35.9.677>
- Rotello, C. M., & Heit, E. (2000). Associative recognition: A case of recall-to-reject processing. *Memory & Cognition*, 28, 907–922.
- Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—retrieving effectively from memory. *Psychonomic Bulletin & Review*, 4, 145–166. <https://doi.org/10.3758/BF03209391>
- Smith, S. M. (1979). Remembering in and out of context. *Journal of Experimental Psychology: Human Learning and Memory*, 5, 460–471. <https://doi.org/10.1037/0278-7393.5.5.460>
- Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, 31, 137–149. <https://doi.org/10.3758/BF03207704>
- Steyvers, M., & Malmberg, K. J. (2003). The effect of normative context variability on recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 760–766. <https://doi.org/10.1037/0278-7393.29.5.760>
- Stretch, V., & Wixted, J. T. (1998). On the difference between strength-based and frequency-based mirror effects in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 1379–1396. <https://doi.org/10.1037/0278-7393.24.6.1379>
- Tversky, B. (1973). Encoding processes in recognition and recall. *Cognitive Psychology*, 5, 275–287.

- Watkins, M. J., LeCompte, D. C., & Kim, K. (2000). Role of study strategy in recall of mixed lists of common and rare words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 239–245.
- Wagenmakers, E.-J., Krypotos, A.-M., Criss, A. H., & Iverson, G. J. (2012). On the interpretation of removable interactions: A survey of the field 33 years after Loftus. *Memory & Cognition*, *40*, 145–160. <https://doi.org/10.3758/s13421-011-0158-0>
- Wixted, J. T. (1992). Subjective memorability and the mirror effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 681–690.
- Yonelinas, A. P. (1997). Recognition memory ROCs for item and associative information: The contribution of recollection and familiarity. *Memory & Cognition*, *25*, 747–763. <https://doi.org/10.3758/BF03211318>
- Zechmeister, E. B., Curt, C., & Sebastian, J. A. (1978). Errors in a recognition memory task are a U-shaped function of word frequency. *Bulletin of the Psychonomic Society*, *11*, 371–373.