

# An Observation on the Role of Context Variability in Free Recall

Jason L. Hicks  
Louisiana State University

Richard L. Marsh and Gabriel I. Cook  
University of Georgia

The authors conducted 3 experiments investigating the effect of context variability and word frequency on free recall. Context variability refers to the number of preexperimental contexts in which a given word is experienced. Both between-subjects and within-subjects manipulations of context variability demonstrated a distinct advantage for low context variability words. Standard word frequency effects were obtained in 2 of the 3 experiments, but the common finding of no word frequency differences in mixed lists of high and low word frequency may depend on the level (low vs. high) of context variability. The authors speculate that the advantage for low context variability items may accrue from better item-to-list context associations or better storage of contextual information as a consequence of the smaller preexperimental contextual fan that these items possess.

*Keywords:* context variability, free recall, memory, word frequency

Context variability as defined here refers to how many preexperimental contexts are associated with a written word. Steyvers and Malmberg (2003) defined high versus low context variability by counting the number of separate texts (on topics such as business, health, language, arts, etc.) in which a word appeared in the Touchstone Applied Science Associates corpus. Low context variability items such as *plasma* (low word frequency) and *planet* (high frequency) occur in many fewer semantic contexts than do high context variability items such as *prayer* (low frequency) and *meal* (high frequency). Thus, the frequent occurrence of *planet* is associated with just a few contexts such as astronomy whereas the frequent word *meal* occurs in a multitude of everyday contexts. Steyvers and Malmberg demonstrated that low context variability confers a recognition memory advantage for words as compared with high context variability items, much like the effect of normative word frequency. Moreover, context variability produces the classic mirror effect, such that low context variability items have a higher hit rate, whereas high context variability items produce a much larger false alarm rate. Cook, Marsh, and Hicks (in press) discovered that like many mirror effects in recognition, the hit rate advantage is due to recollection, and the higher false alarm rate for high context variability items is due to familiarity (cf. Arndt & Reder, 2002; Hirshman, Fisher, Henthorn, Arndt, & Passannante, 2002; Joordens & Hockley, 2000). In this brief report, we investigated the influence of context variability (and word frequency) on free recall.

There are a number of reasons why such an investigation merits attention, but perhaps none is more important than the fact that

contextual information is emerging as a critical component in recent theoretical models of memory. For example, Dennis and Humphreys's (2001) Bind, Cue, and Decide Model of Memory (BCDMEM model) is based entirely on interference to memory accruing not from other list items, but rather from contextual elements stored at the time of encoding and those from a preexisting exposure to the real world. When evaluating a candidate memory, the retrieved contextual information is compared with a composite vector representing the entire list context. In another approach, Howard and Kahana (1999, 2002) have argued that a candidate memory evokes retrieval of its temporal context in the study list (temporal context model, or TCM). They argued further that this retrieved contextual information can serve as a retrieval cue for other list items. Moreover, in that theory, the time-of-test context also plays a role in the ability of a rememberer to gain access to memories of the original context, as it does in the BCDMEM model more generally. Thus, unlike older models of memory like the Search of Associative Memory (SAM) model and its progeny that focus more on inter-item associations, these newer models have highlighted the importance that contextual information has in retrieval tasks. Of course, inter-item associations are still important in a model like TCM, but the emphasis has shifted to contextual information.

We considered two alternative predictions; the first is described in this paragraph, and the second is described in the following one. One hypothesis was that context variability acts in free recall just like word frequency because the effects on recognition memory are similar (e.g., Cook et al., in press). In free recall, pure lists of high frequency words are remembered better than pure lists of low frequency words (e.g., DeLosh & McDaniel, 1996; Duncan, 1974; Gregg, 1976). When both classes of items are intermixed, items of low word frequency are either recalled at rates equivalent to high frequency words or sometimes even better (e.g., Duncan, 1974; Gregg, Montgomery, & Castaño, 1980). Of course, such outcomes are generalizations because qualifications arise, such as when very long lists are used (e.g., Balota & Neely, 1980; McDaniel, DeLosh, & Merritt, 2000). Nonetheless, better recall of high frequency words has been explained by the notion that they supported more

---

Jason L. Hicks, Department of Psychology, Louisiana State University; Richard L. Marsh and Gabriel I. Cook, Department of Psychology, University of Georgia.

Gabriel I. Cook is now at the Department of Psychology, Claremont McKenna College.

We thank Dean Stith for his dedicated help in collecting the data.

Correspondence concerning this article should be addressed to Richard L. Marsh, Department of Psychology, University of Georgia, Athens, GA 30602-3013. E-mail: rmarsh@uga.edu

inter-item associations and that presenting low frequency words in a mixed list disrupts the formation of such associations. If context variability has the same effect as word frequency, then high context variability items would support greater free recall in pure lists than low context variability items. Moreover, such an effect might be removed in mixed lists of high and low context variability items.

We then considered a different, second alternative. Because free recall depends on generating and testing candidate memories (e.g., Anderson & Bower, 1972; Tulving, 1976), inter-item associations are only one means of generating candidates. Another important means is to use the item-to-studied list context association (Anderson & Bower, 1972). This kind of association is from a studied word to environmental and spatiotemporal contextual elements occurring at the time a word is experienced. We might assume that high and low context variability items do not differ in their ability to support inter-item associations, but they may differ in the degree to which they are bound to the studied list context. Low context variability items have fewer preexisting associations than do high context variability items. Therefore, it may be easier to associate low context variability items with the experimental context on the basis of theories advocating fan effects in memory because they have a smaller preexperimental fan (e.g., Anderson & Reder, 1999). Accordingly, associating the experimental context to an item that has fewer preexisting associations is easier. Howard and Kahana's (2002) TCM may predict the same because recovery of the temporal context for low context variability items might be better than that for high context variability items if the association to the experimental context is stronger. By both accounts, recall of low context variability items should be advantaged.

Against this backdrop, we conducted three free recall experiments to investigate which of these two alternative predictions would be borne out. In Experiment 1, high and low word frequency and high and low context variability were crossed, and the resulting four types of lists were presented to different participants in an entirely between-subjects design. In Experiments 2 and 3, word lists either of high or low word frequency had intermingled items of different context variabilities, or pure lists of high and low context variability had intermingled words of different word frequencies. For brevity, Experiments 2 and 3 are presented together.

### Experiment 1

Experiment 1 was designed to ascertain whether context variability affected free recall. One prediction was that context variability acts like word frequency; thus, lists of high context variability items should be recalled better than lists of low context variability items. The alternative prediction argues that low context variability items have fewer preexisting contextual associations (i.e., a smaller preexperimental fan), which may facilitate forming stronger item-to-context associations; thus, low context variability items are recalled more frequently.

### Method

**Participants.** Undergraduate students from the University of Georgia volunteered in exchange for partial credit toward a course research requirement. Each participant was tested individually in a session that lasted approximately 20 min. We had intended on testing 25 participants in each of the four between-subjects conditions that result from crossing high and

low context variability with high and low word frequency. Accidentally, an extra participant was tested in the low context variability and high word frequency condition. Thus, a total of 101 people participated in this experiment. Assignment to the four conditions was on a continuous rotation schedule.

**Materials and procedure.** Each of the four item classes (low-low, high-low, low-high, and high-high) had 40 items that were drawn from the appendix of Steyvers and Malmberg's (2003) report. We used their classification of assignment of items to classes and also chose as many items as we could that did not have obvious semantic associations to one another. Software was written that presented each of the 40 items individually in the center of the computer monitor for 5 s each. The order of the studied items was randomized anew for each participant tested. Participants were asked to study the words for an unspecified memory test. After the learning session concluded, an instruction screen came up that indicated that the experimenter had a paper-and-pencil task to be performed. This distractor task was working on mazes for 3 min as timed by the computer. Following this task, free recall was assessed. Participants wrote down as many words as they could remember studying in the 4-min recall period (also timed by the computer).

### Results and Discussion

Unless an associated  $p$  is specifically provided with a statistical test, the probability of a type I error did not exceed the conventional 5% in this and the two subsequent experiments. Free recall was scored by the experimenter against master lists for each of the four conditions. We did not report intrusion errors because of their rarity in free recall (e.g., Bower & Mann, 1992). Table 1 summarizes the average proportion of the 40 studied words that were recalled. We conducted a 2 (high vs. low context variability)  $\times$  2 (high vs. low word frequency) analysis of variance (ANOVA) model on the data. Low context variability items were free-recalled more often than were high context variability items,  $F(1, 97) = 11.31$ . In addition, high frequency words were better recalled than those of low frequency,  $F(1, 97) = 7.20$ . The interaction was not statistically significant ( $p > .20$ ). Because context variability is a relatively new variable, simple effects analyses were conducted to test whether context variability led to different levels of free recall at each level of word frequency (i.e., within the rows of Table 1).

Table 1  
Average Free Recall as a Function of Context Variability and Word Frequency in Experiments 1–3

Experiment and list type	Context variability		
	Low	High	Average
Experiment 1			
Low word frequency	.28 (.02)	.24 (.02)	.26 (.01)
High word frequency	.35 (.02)	.27 (.02)	.31 (.02)
Average	.32 (.01)	.25 (.01)	
Experiment 2			
Low word frequency	.30 (.02)	.19 (.02)	.25 (.02)
High word frequency	.36 (.02)	.29 (.02)	.33 (.02)
Average	.33 (.01)	.24 (.01)	
Experiment 3			
Low word frequency	.26 (.03)	.22 (.02)	.24 (.02)
High word frequency	.34 (.03)	.22 (.02)	.28 (.01)
Average	.30 (.02)	.22 (.02)	

*Note.* Standard errors are in parentheses.

With entirely low frequency words, low context variability items were recalled numerically more often than high context variability items, although the effect fell on the margin,  $t(48) = 1.79$ ,  $p = .07$ . With high frequency words, items experienced in fewer preexperimental contexts were recalled more often than items experienced in a greater variety of contexts,  $t(48) = 2.87$ . The finding that high frequency words were recalled more often as compared with low frequency words merely replicates what has been reported for many years.

The novel finding was that words having fewer preexperimental contextual associations were recalled more often than those with many such associations (at both levels of word frequency). Obviously, context variability is dissociated from word frequency with a free-recall task. At this juncture, one could advocate a mechanism by which items with low context variability have fewer associations, allowing them to be better bound to the studied list context. In other words, their preexperimental fan is smaller, thereby allowing the association to the list context to be more effective at retrieval. On the other hand, nothing in the data contravene an interpretation that the same smaller fan does not also facilitate more effective inter-item associations. Experiments 2 and 3 were conducted to partially address that issue.

### Experiments 2 and 3

In the word frequency literature, researchers have reported that mixed lists of high and low word frequency items usually remove the high word frequency advantage in recall. The presence of the low frequency items may attract some attention and reduce the number or quality of inter-item associations that are formed for high frequency words (e.g., Duncan, 1974; Gregg et al., 1980). If the same logic is applied to context variability, mixed lists of high and low context variability items may reduce or eliminate the low context variability advantage. On the other hand, if the advantage is preserved, then the locus of the effect can be more strongly linked to item-to-list context associations. Consequently, for Experiment 2, we mixed high and low context variability items together in a single list. Given that the simple effects may have been different in Experiment 1 as a function of word frequency, we tested both a low and a high word frequency condition (with context variability manipulated within lists). The purpose of Experiment 3 was somewhat more ancillary, but we wanted to ascertain if context variability affected whether the mixed-list absence of the word frequency effect would be obtained. Stated alternatively, there should have been no word frequency differences in mixed lists of high and low word frequency items, but we wanted to determine if equivalence held at both levels of context variability.

### Method

*Participants.* Undergraduates from the University of Georgia volunteered in exchange for partial credit toward a course requirement. Each participant was tested individually in a session that lasted approximately 20 min. The distribution of the 111 participants tested was as follows: In Experiment 2, 28 people were tested with the low frequency list, and 29 were tested with the high frequency list. In Experiment 3, 27 people were tested in each condition.

*Materials and procedure.* Experiments 2 and 3 were virtually identical to Experiment 1 in all of their essential properties except for the compo-

sition of the 40-item study list. In Experiment 2, two groups of participants were tested, one with high frequency words and one with low frequency words. Within a given list, 20 words were low context variability items and 20 were high context variability items. As in Experiment 1, the order of presentation was randomized anew for each participant tested, and this randomization was done without regard to context variability. In Experiment 3, two groups of participants were also tested; one group studied only high context variability items, whereas the other group studied only low context variability items. Within each list, half of the words were low frequency and half were high frequency. In all other respects, these experiments were identical to Experiment 1.

### Results and Discussion

The results of Experiment 2 and 3 are summarized in Table 1 as proportions of studied items that were recalled. We conducted a 2 (high vs. low context variability)  $\times$  2 (high vs. low word frequency) mixed-model ANOVA on the data from Experiment 2. Low context variability items were more often recalled than high context items,  $F(1, 55) = 23.97$ . Because word frequency was manipulated between subjects, we obtained the standard finding that high frequency items were recalled more often than low frequency items,  $F(1, 55) = 11.25$ . Context variability and word frequency did not statistically interact. Simple effects analyses were conducted within each word frequency condition (within the rows of Table 1). Free recall of low context items was better than that of high context items at both low word frequency,  $t(27) = 4.32$ , and high word frequency,  $t(28) = 2.71$ . The replication of the advantage of low context variability items in a within-subjects design suggests that inter-item associations probably do not differ across levels of context variability or that presenting items that have been experienced in many preexperimental contexts does not affect the formation or use of these associations. Consequently, these data may favor an explanation more in line with low context variability items forming better item-to-context associations.

The results of Experiment 3 were analyzed in an analogous mixed-model ANOVA to Experiment 2. Items having low context variability once again led to better free recall than items with high context variability,  $F(1, 52) = 7.42$ . Although the word frequency effect should have been eliminated because the lists were of mixed word frequency, there was a residual effect favoring better recall of high frequency words,  $F(1, 52) = 3.44$ ,  $p = .06$ , and a marginal interaction,  $F(1, 52) = 3.97$ ,  $p = .06$ . The simple effects analysis (within the columns of Table 1) showed that the high word frequency advantage was preserved for low context variability items,  $t(26) = 2.42$  but was eliminated for high context variability items,  $t(26) < 1$ . One way to interpret the preserved word frequency effect is that low context variability lists placed participants in a slightly higher range of the response scale (by approximately 8%). Because participants were better able to recall items and because items did have associations to other list members (high frequency items in particular), the high word frequency advantage may have been preserved only because participants were better able to recall items initially. We admit that this explanation warrants further scrutiny, but for now, it is one viable hypothesis for the interaction. We turn now to considering these results more generally.

## General Discussion

The purpose of this study was to determine whether context variability affected free recall. The unequivocal answer to that question is that low context variability confers an advantage to recall from long-term memory. That finding also appears to be rather robust because it survives both between- and within-subjects manipulations, and it appears generally to be found at both levels of word frequency (albeit marginal in Experiment 1 with low word frequency), and the sample sizes used here were only moderate. Given that low context variability confers a recognition memory advantage, (Cook et al., in press; Steyvers & Malmberg, 2003), this variable has a similar outcome on both retrieval processes (i.e., recall and recognition). The obvious difference between items of high versus low context variability is the number of preexperimental contextual associations they possess. Any theoretical analysis of these novel findings may hinge critically on this difference between the two types of items.<sup>1</sup>

One a priori hypothesis could have been that multiple associations to different contexts would assist in forming associations to other items, thereby giving high context variability items an advantage in free recall. On the other hand, according to newer theories that incorporate the fan effect into their modeling (e.g., Reder et al., 2000), few contextual associations may assist in reinstating either inter-item or list-to-context associations during retrieval. Although Experiments 2 and 3 by no means completely rule out that numbers of inter-item associations may differ across levels of contextual variability, intermixing different levels did not appear to change performance at all as it usually does for word frequency. Therefore, our preferred argument is that low context variability items have stronger item-to-list context associations and when the episodic context of study is used as a retrieval cue, more of these items can be recalled. Of course, this is a simplistic explanation based on older generate-and-test models of free recall that used the contextual marker "on most recent list" as a cue to free recall (e.g., Anderson & Bower, 1972).

Proponents of more recent models such as TCM would argue that the context of a studied item is retrieved when it is recalled and that contextual information is used to help retrieve the next item in the participant's recall protocol (e.g., Howard & Kahana, 1999, 2002). To integrate the present findings into that model, one would have to assume that the smaller preexperimental fan of low context variability items aids in storage of contextual information at the time the word is studied. When cued by the contextual information of the previous item, an item possessing that greater contextual information stored in memory has a higher probability of being retrieved. The present results are also consistent with the BCDMEM model (Dennis & Humphreys, 2001),<sup>2</sup> which uses a similar assumption of better encoding of contextual information. In that model, candidates are evaluated by comparing contextual information from their episodic and preexperimental exposures to a context vector representing the entire study list. Using the present argument, we would contend that the quality of contextual sampling or the quality of that vector may be greater when studied items have few preexperimental associations as opposed to many. Thus, even a within-subjects manipulation of context variability would improve the information stored in that vector itself. Thus, there seems to be some straightforward ways to incorporate the current findings into newer models of memory, especially when

the editing process (i.e., recognition-like process) of free recall is factored into the processes that are needed for free recall. As this variable called context variability is understood more clearly through additional research, one challenge will be to assess how the current data can be used to discriminate among the finer grained predictions of some newer models.

As a final consideration, we note that context variability and word frequency are naturally quite highly correlated (Dennis & Humphreys, 2001). Thus, until Steyvers and Malmberg's (2003) objective approach of disentangling context variability and word frequency, the vast majority of the work reporting word frequency effects over the past half century has undoubtedly confounded context variability with word frequency. Thus, generalizations made during this period, such as the one that we stated in the introduction about high frequency words being better recalled, are all based on materials that are likely to have been confounded with context variability. The interaction we found in Experiment 3 must raise eyebrows, given that word frequency had different effects at different levels of context variability. In addition, both recognition and recall are better with low context variability items, whereas word frequency is dissociated (recognition is better with low word frequency, and recall is better with high word frequency). However, context variability and word frequency rarely interacted in our recognition memory work (Cook et al., in press) and did not interact in Experiments 1 and 2. From that perspective, the confounding of these two variables over the years may have only helped the word frequency effects in recognition memory. The most seriously compromised studies would be those that directly compared recognition to free recall in which the context variability and word frequency dimensions could have been at direct odds with one another in free recall. Thus, our analysis suggests that some past recognition results actually might not be as robust as previously believed, and some free recall results may require some serious reevaluation.

<sup>1</sup> A reviewer noted that low context variability items have greater average concreteness and imaginability values. Although this is true, we believe that words are rated as more concrete because they have fewer context associations. Using an objective environmental statistic such as context variability, we have circumvented the subjective evaluation of what makes a concept more or less concrete. Thus, Steyvers and Malmberg's (2003) corpus is not confounded by concreteness but rather suggests a fundamental reason why participants give the concreteness ratings that they do.

<sup>2</sup> Technically, the BCDMEM model has not been officially extended to free recall as of this writing. Thus, our remarks should be taken as speculation only.

## References

- Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review*, 79, 97–123.
- Anderson, J. R., & Reder, L. M. (1999). The fan effect: New results and new theories. *Journal of Experimental Psychology: General*, 128, 186–197.
- Arndt, J., & Reder, L. M. (2002). Word frequency and receiver operating characteristic curves in recognition memory: Evidence for a dual-process interpretation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 830–842.

- Balota, D. A., & Neely, J. H. (1980). Test-expectancy and word-frequency effects in recall and recognition. *Journal of Experimental Psychology: Human Learning & Memory*, 6, 576–587.
- Bower, G. H., & Mann, T. (1992). Improving recall by recoding interfering material at the time of recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 1310–1320.
- Cook, G. I., Marsh, R. L., & Hicks, J. L. (in press). The role of recollection and familiarity in the context variability mirror effect. *Memory & Cognition*.
- 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.
- Dennis, S., & Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review*, 108, 452–478.
- Duncan, C. P. (1974). Retrieval of low-frequency words from fixed lists. *Bulletin of the Psychonomic Society*, 4, 137–138.
- Gregg, V. H. (1976). Word frequency, recognition and recall. In J. Brown (Ed.), *Recall and recognition* (pp. 183–216). Oxford, England: Wiley.
- Gregg, V. H., Montgomery, D. C., & Castaño, D. (1980). Recall of common and uncommon words from pure and mixed lists. *Journal of Verbal Learning & Verbal Behavior*, 19, 240–245.
- Hirshman, E., Fisher, J., Henthorn, T., Arndt, J., & Passannante, A. (2002). Midazolam amnesia and dual-process models of the word-frequency mirror effect. *Journal of Memory & Language*, 47, 499–516.
- Howard, M. W., & Kahana, M. J. (1999). Contextual variability and serial position effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 923–941.
- Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, 46, 269–299.
- Joordens, S., & Hockley, W. E. (2000). Recollection and familiarity through the looking glass: When old does not mirror new. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 1534–1555.
- McDaniel, M. A., DeLosh, E. L., & Merritt, P. S. (2000). Order information and retrieval distinctiveness: Recall of common versus bizarre material. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 1045–1056.
- 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.
- Steyvers, M., & Malmberg, K. J. (2003). The effect of normative context variability of recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 760–766.
- Tulving, E. (1976). Ecphoric processes in recall. In J. Brown (Ed.), *Recall and recognition* (pp. 37–73). Oxford, England: Wiley.

Received January 5, 2005

Revision received March 11, 2005

Accepted March 15, 2005 ■

### Correction to Sahakyan and Delaney (2005)

The article “Directed Forgetting in Incidental Learning and Recognition Testing: Support for a Two-Factor Account,” by Lili Sahakyan and Peter F. Delaney (*Journal of Experimental Psychology: Learning, Memory, and Cognition*, Vol. 31, No. 4, pp. 789–801), was misidentified in the July issue as an Observation. The manuscript was accepted as a Regular Article.