Journal of Memory and Language 64 (2011) 119–132

Contents lists available at ScienceDirect

Journal of Memory and Language

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

The effects of word frequency and context variability in cued recall

Amy H. Criss *, William R. Aue, Larissa Smith

Department of Psychology, Syracuse University, United States

ARTICLE INFO

Article history:
Received 22 March 2010
Revision received 14 October 2010
Available online 16 November 2010

Keywords:
Cued recall
Episodic memory
Word frequency
Context variability
Mathematical models
Memory models

ABSTRACT

Normative word frequency and context variability affect memory in a range of episodic memory tasks and place constraints on theoretical development. In four experiments, we independently manipulated the word frequency and context variability of the targets (to-be-generated items) and cues in a cued recall paradigm. We found that high frequency targets were better recalled in both pure and mixed lists, even when context variability was held constant. High frequency cues were slightly more effective, but this benefit was eliminated when context variability was held constant. Low context variability cues were most effective while the context variability of the target had little effect on performance. The data suggest that words with fewer pre-experimental connections are better able to isolate the list and that generation of an item from memory benefits from frequency, perhaps due to the ease of generating common orthographic and phonological features. Implications for current models of memory and the prospects of future models are discussed.

Introduction

Properties of to-be-remembered items partially determine the results of later memory tasks. For example, studies have shown differences in memory between pictures vs. words, words vs. non-words, emotional vs. neutral items, high vs. low arousal items, etc. (e.g., Bradley, Greenwald, Petry, & Lang, 1992; Gillund & Shiffrin, 1981; Greene, 2004; Grider & Malmberg, 2008; Kapucu, Rotello, Ready, & Seidl, 2008; Nelson, Reed, & McEvoy, 1977; Onyper, Zhang, & Howard, 2010; Paivio, 1971; Snodgrass & McClure, 1975). Our goal here is to evaluate item properties (i.e., word frequency and context variability) that play a role in the successful generation of a target word and the effectiveness of a cue word.

Word frequency

Normative word frequency (WF) is one property of words that has received much empirical and theoretical attention. In single item recognition, uncommon low frequency (LF) words are remembered better than common high frequency (HF) words. Typically this manifests as a mirror pattern where hit rates (HR) are higher and false alarm rates (FAR) are lower for LF than HF words (e.g., Glanzer & Adams, 1985; Schulman, 1967). This word frequency mirror effect is a benchmark finding that is accounted for by most models of recognition memory, albeit with different underlying mechanisms.

Critically, the pattern of accuracy for HF and LF words changes when the paradigm by which memory is evaluated changes. In a free recall task where participants are asked to generate as many target words from the study list as possible without being provided any explicit memory cue, more HF than LF words are successfully recalled (DeLosh & McDaniel, 1996; Gregg, 1976; Hall, 1954). This pattern holds when the study list is composed of a single frequency (either all HF or all LF). The HF benefit in recall is less reliable, sometimes absent, or even reversed when

* Corresponding author. Address: Department of Psychology, Syracuse University, Syracuse, NY 13244, United States.

E-mail address: acriss@syr.edu (A.H. Criss).

0749-596X/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jml.2010.10.001

the encoded list is mixed in WF composition, containing both HF and LF words. This is referred to as the mixed list paradox (Gillund & Shiffrin, 1984; Gregg, Montgomery, & Castaño, 1980; May, Cuddy, & Norton, 1979; Watkins, LeCompte, & Kim, 2000). The mixed list paradox is also present in immediate serial recall (Hulme, Stuart, Brown, & Morin, 2003). This has been interpreted as evidence that pre-experimental associations (assumed to be more plentiful for HF pairs) play an important role in a recall task. This and other dissociations between free recall and single item recognition are difficult to account for with a single model. Such dissociations have contributed to the current state of the field where models tend to be applied to either single item recognition or free recall, but not both.

Context variability

Nearly all studies of WF confound WF with context variability (CV)due, in part, to the high correlation between the two measures (Dennis & Humphreys, 2001; Steyvers & Malmberg, 2003). CV is a measure of the number of *different* contexts in which a word appears in a corpus. Consider the words *soccer* and *wrist*, both low frequency words. *Soccer* almost always appears in the context of sports while *wrist* may appear in a variety of contexts such as anatomy, medicine, jewelry, sports, etc. Soccer is an example of a low CV (LCV) word and wrist is an example of a high CV (HCV) word. Steyvers and Malmberg (2003) were among the first to empirically address the confound between WF and CV in episodic memory. They did so by constructing stimulus sets where the mean WF was approximately the same for both the HCV and LCV word sets and the mean CV was approximately the same for both the HF and LF word sets (this same stimulus set is used in Experiments 3 and 4 of the current paper, see Table 3). In other words, they orthogonally manipulated WF and CV. In their single item recognition experiment, Steyvers & Malmberg found an advantage for LF and LCV words in the form of simultaneous mirror effects for both WF and CV. In experiments where subjects were asked to provide a subjective report of their recognition decisions, LCV words had higher reports of recollection in the HRs and HCV words had higher reports of familiarity in the FARs; the same pattern has also been observed for WF (Cook, Marsh, & Hicks, 2006). To summarize, in single item recognition experiments, independent manipulations of WF and CV (e.g., not confounded with one another) affect performance in the same way. Specifically, LCV and LF words are better recognized than HCV and HF words, respectively.

The effects of WF and CV do not show the same pattern in a free recall task. Recall that pure lists result in a HF advantage in free recall (e.g., Gregg, 1976; Hall, 1954). Both between- and within-subject manipulations of CV reveal better performance for LCV compared to HCV words regardless of WF (Hicks, Marsh, & Cook, 2005). The LCV advantage in free recall holds for the same stimulus set used by Steyvers and Malmberg (2003) and when the stimulus set is further constrained so that concreteness is controlled (Marsh, Meeks, Hicks, Cook, & Clark-Foos, 2006).

The overall pattern across these studies shows separate and independent effects of CV and WF on episodic memory performance. In recognition, LF and LCV targets are better remembered and LF and LCV foils more likely to be rejected than HF or HCV counterparts. In free recall, LCV and HF words are more likely to be recalled than HCV or LF words, respectively.

Models of single item recognition

Mathematical models that account for the word frequency mirror effect in single item recognition are plentiful, as are the proposed underlying mechanisms (e.g., Glanzer & Adams, 1990; Dennis & Humphreys, 2001; McClelland & Chappell, 1998; Shiffrin & Steyvers, 1997; Reder et al., 2000). We will consider two examples that are most relevant for this work, acknowledging that there are several alternatives. First, consider the Retrieving Effectively from Memory model (REM; Shiffrin & Steyvers, 1997) which attributes the WF mirror effect to the high diagnosticity provided by uncommon features of LF words. For example, LF words are composed of atypical letters and letter combinations relative to HF words (e.g., Cleary, Morris, & Langley, 2007; Criss & Malmberg, 2008; Freeman, Heathcote, Chalmers, & Hockley, 2010; Landauer & Streeter, 1973; Malmberg & Nelson, 2003; Malmberg, Steyvers, Stephens, & Shiffrin, 2002; Zechmeister, 1969). Uncommon words are composed of uncommon features in REM and therefore tend to not match other words by chance, resulting in a lower FAR for LF words. However, matching an uncommon feature during retrieval provides more evidence in favor of that item than does matching a common feature, leading to a higher HR for LF words. In other words, on average LF foils are a poor match to *other* words stored in memory reducing the FAR and LF targets are a good match to their own memory trace increasing the HR.

Second, consider models that attribute the WF effect to interference caused by the large number and variability of prior contexts in which HF words have been encountered (Dennis & Humphreys, 2001; Reder et al., 2000). The Bind Cue Decide Model of Episodic Memory (BCDMEM; Dennis & Humphreys, 2001) operates via a single process where the features of the reinstated study context are matched against all prior contexts in which the test item had been encountered. According to BCDMEM, HF words tend to be experienced in more pre-experimental contexts and thus have more interference and lower accuracy compared to LF words. In the Source of Activation Confusion (SAC; Reder et al., 2000) model, HF words have higher baseline familiarity at the concept node due to the larger number of times they have been previously encountered, resulting in a higher FAR for HF words. Further, LF words are better recollected due to the relatively smaller number of prior contexts in which they appeared. This LF benefit in recollection overcomes the higher baseline familiarity for HF words, resulting in a higher HR for LF words. Thus, both the SAC and BCDMEM models predict that items with many pre-experimental associations are more difficult to remember because the other contexts to which they are associated interfere with remembering the association of the item and the experimental context.

Models of free recall

In contrast to single item recognition, few computational models of free recall account for the WF effect. The Search of Associative Memory (SAM; Raaijmakers & Shiffrin, 1981) model is one model that accounts for the word frequency effect in both recall and recognition. SAM assumes that HF words have stronger associative connections to other items. HF words benefit from stronger associations developed prior to the experiment and stronger associations developed between items within the experiment. Thus, HF words serve as more effective cues to access any other item than do LF words (Gillund & Shiffrin, 1984). Unfortunately, SAM is generally not considered a viable model for single item recognition owing to several key pieces of data (e.g., Ratcliff, Clark, & Shiffrin, 1990; Shiffrin, Ratcliff, & Clark, 1990) which led to the development of a new generation of Bayesian models (i.e., Dennis & Humphreys, 2001; McClelland & Chappell, 1998; Shiffrin & Steyvers, 1997). Nevertheless, the associative hypothesis for the HF benefit in free recall persists and is frequently evoked to explain the WF effect in free recall (e.g., Ozubko & Joordens, 2007; Stuart & Hulme, 2000). Two other dominant hypotheses about the WF effect in free recall are the order-encoding hypothesis (DeLosh & McDaniel, 1996; Merritt, DeLosh, & McDaniel, 2006; Toglia & Kimble, 1976) and the recency hypothesis (Tan & Ward, 2000; Ward, Woodward, Stevens, & Stinson, 2003). According to the order-encoding hypothesis, unusual items, including LF words, capture attention in service of encoding the features of the item itself and at the cost of encoding information about the order of the item in the encoded list. The recency hypothesis attributes better memory for HF words to the more frequent and more recent (relative to test) rehearsal of HF words within a study session. When recency and number of rehearsals is equated, HF words are still better recalled, thus the recency hypothesis also incorporates the assumption that HF words benefit from inter-item associations more than LF words.

If the goal is to have a comprehensive model of episodic memory, then the field is in an unsatisfactory state. There are several well-specified models of free recall that are supported by empirical data. There are also several models of recognition that are supported by empirical data. However, there exists no satisfactory model of both tasks that accounts for empirical dissociations between the tasks due to variables such as word frequency. One broad goal of this research program is to begin considering how to explain behavior in both free recall and single item recognition and eliminate the theoretical gap between models of these tasks. The specific strategy in this paper is to conduct cued recall experiments, a task that shares properties with both single item recognition and free recall.

In cued recall, participants receive a cue at test which they use to probe memory for a specific item. The cue may be a category label, an item related to the target, or an item from the experimental list, among other possibilities. The most common method for cued recall and the one used here involves presentation of pairs during encoding, one of which is used as a cue and the other as the target in a later memory test. This cued recall paradigm shares properties with both single item recognition and free recall: participants are provided with a cue (as in single item recognition) and are asked to generate the target (as in free recall) that was paired with the cue at study. Many of the models and theories described above have not been directly applied to cued recall and therefore it is not possible to generate exact predictions. Throughout this paper, we assume a simple model for cued recall based on models of cued recall that have been implemented within the SAM and REM frameworks (e.g., Gillund & Shiffrin, 1984; Diller, Nobel, & Shiffrin, 2001). The presented cue is compared to the episodic memory traces and a match to each memory trace is computed just as in single item recognition (in which the decision is based on the combined information about those matches) and thus our hypotheses about the role of the cue follows from recognition data. After the matching process, a single memory trace is sampled (chosen in proportion to how well it matches the cue) from memory. The participant then attempts to retrieve the target contained in the sampled trace, as in free recall and thus hypotheses about the role of the target are based on free recall data.

In this paper, we will evaluate properties of items that contribute to the effectiveness of a cue and successful generation of a target by independently manipulating properties of the cue and target in a cued recall paradigm. We will consider the item properties of WF and CV because they each play a central role in different recognition memory models and because the effects of each variable in single item recognition and free recall are well documented. A better understanding of the role of WF and CV will provide strong constraints on models as the field moves toward integrating models of free recall and single item recognition.

General method

Participants

All participants were from Syracuse University and participated in exchange for $10 per hour or for a class requirement.

Design

The experiment consisted of blocks each containing 20 study trials, a 60 s distracter task, a 20 trial test list, and a 90 s break, in that order. The study-distracter-test-break cycle repeated multiple times with no items repeating across blocks. During the distracter task participants kept a running summation of a series of individual digits. The break between each study-test block was intended to reduce fatigue and interference between blocks.

During study, pairs of items were presented for 3 s after which participants made a judgment about the pair. Participants were not aware which item of each pair would later be the to-be-generated target and which would serve as the cue for that target. The words were presented side-by-side during encoding and the screen position of the target and cue were randomly selected on each trial. Encoding task was manipulated between subjects. Participants

Table 1
Mean response probabilities for each condition in Experiment 1 as a function of encoding task. Word frequency of the cues and targets were orthogonally manipulated (HF = high frequency, LF = low frequency) and presented in pure lists during encoding. Standard errors are in parentheses.

	HF target		LF target	
	HF cue	LF cue	HF cue	LF cue
Experiment 1: Sentence task				
P (correct)	0.324 (0.040)	0.298 (0.037)	0.217 (0.031)	0.170 (0.029)
P (intrusion)	0.104 (0.017)	0.091 (0.013)	0.076 (0.014)	0.098 (0.016)
P (no response)	0.572 (0.042)	0.611 (0.041)	0.707 (0.036)	0.731 (0.036)
Experiment 1: Association task				
P (correct)	0.384 (0.038)	0.297 (0.036)	0.200 (0.030)	0.147 (0.028)
P (intrusion)	0.093 (0.017)	0.069 (0.013)	0.088 (0.013)	0.097 (0.015)
P (no response)	0.522 (0.041)	0.634 (0.040)	0.712 (0.035)	0.757 (0.035)

Table 2
Mean response probabilities for each condition in Experiment 2 as a function of encoding task. Word frequency of the cues and targets were orthogonally manipulated (HF = high frequency, LF = low frequency) and presented in mixed lists during encoding. Standard errors are in parentheses.

	HF target		LF target	
	HF cue	LF cue	HF cue	LF cue
Experiment 2: Sentence task				
P (correct)	0.412 (0.036)	0.343 (0.031)	0.248 (0.027)	0.238 (0.028)
P (intrusion)	0.088 (0.016)	0.110 (0.016)	0.102 (0.017)	0.095 (0.022)
P (no response)	0.500 (0.036)	0.547 (0.034)	0.650 (0.027)	0.667 (0.032)
Experiment 2: Association task				
P (correct)	0.373 (0.036)	0.297 (0.030)	0.193 (0.027)	0.148 (0.027)
P (intrusion)	0.097 (0.016)	0.097 (0.016)	0.087 (0.016)	0.140 (0.022)
P (no response)	0.530 (0.036)	0.607 (0.033)	0.720 (0.026)	0.712 (0.032)

either generated a sentence about the pair of items and rated the difficulty of doing so (on a 9-point scale where 1 – *very easy* and 9 – *very difficult*), or judged the degree of association between the pair of items (on a 9-point scale where 1 – *not at all associated* and 9 – *highly associated*). Two different tasks were used to alleviate any concerns that a sentence generation task may have interacted with the experimenter assigned cues and targets.[1]

The cued recall test was self-paced with the order of cues randomized for each participant. On each trial, a cue was presented in the middle of the screen and participants were asked to type the word it was paired with during study. Participants were allowed to provide no answer (and indicated this by typing "no"). Participants were informed that spelling "did not count" to encourage them to respond even for words that may be difficult to spell correctly. For example we did not want participants to withhold LF targets they had generated simply because they were unsure of the spelling.

Each response was scored as correct, an intrusion, or no response. Spelling errors and typographical errors were forgiven and scored appropriately. For example if the participant typed "np" that was scored as no response (e.g., we assume participants intended to type "no"), if participants typed "freind" for the target word "friend" that was scored as correct.[2] For each participant, for each condition, the sum of the three response possibilities must sum to the total number of trials; an increase in correct responses must be accompanied by a decrease in one or both of the error responses. There are many possible types of intrusions (e.g., intra-list, extra-list, extra-experiment, semantically similar to the target, etc.), however, the overall rate of intrusions was low and prevented meaningful analysis of the types of intrusions. The dependent measure of interest is proportion correct, however the proportion of intrusions and failures to respond are included for completeness.

Analysis of encoding task

Readers may have been concerned that the sentence task differentially benefited one class of cues, and so we replicated each experiment with a different encoding task, the association task. Here we evaluate whether encoding task has any effect on accuracy in cued recall. For proportion correct in each of the four experiments, a mixed analysis of variance (ANOVA) was conducted with encoding task included as a between-subject factor. There was no main effect of encoding task and no interaction between encoding task and any other variable in any experiment (all F's < 2.91 and all p's > .094) with one exception. In Experiment 4, there was a marginally significant interaction between encoding task and CV of the target, $F(1, 61) = 3.976$, $p = 0.051$. The interaction appears to stem from a small difference between proportion correct in the two encoding tasks for high CV (sentence task: $M = .360$, $SE = .031$ and association task: $M = .412$, $SE = .033$) but not for low CV targets (sentence task: $M = .380$, $SE = .030$

[1] We thank Simon Dennis and Geoff Ward for this suggestion.

[2] Both strict (only exact matches were coded as correct) and lenient scoring procedures were adopted. Overall accuracy was higher under lenient scoring but the pattern of data did not change. Thus data from the lenient scoring rule are reported.

Table 3
Word frequency values for the stimuli used in Experiments 3 and 4 from 3 different corpora. The reported values are not on the same scale, instead they follow the convention for each database to ensure consistency with published literature. The TASA database has 10,710,325 words and is available at http://psiexp.ss.uci.edu/research/software.htm. The mean word frequency values reported for TASA are based on raw counts for each word. The Google database has over a trillion words and is available at http://mall.psy.ohio-state.edu/wiki/index.php/Main_Page. The mean word frequency values reported for Google are based on counts per million for each word. The HAL database has approximately 131 million words and is available at http://elexicon.wustl.edu/. The mean word frequency values reported for HAL are based on the log transform of the raw frequency counts for each word. (HF = high frequency, LF = low frequency, HCV = high context variability, LCV = low context variability).

Database	Context variability	Word frequency		Mean
		HF	LF	
logHAL	HCV	10.40	8.10	9.25
	LCV	9.59	7.34	8.48
	mean	9.99	7.72	
Google	HCV	71.37	12.15	42.61
	LCV	67.41	6.14	38.37
	mean	69.41	9.29	
TASA	HCV	1160.01	95.29	627.65
	LCV	1224.17	94.56	659.36
	mean	1192.09	94.92	

and association task: $M = .393$, $SE = .032$). Separate ANOVAs were conducted for each encoding task and there was no main effect of CV of the target in either task (sentence task: $F(1, 32) = 3.207$, $p = .083$ and association task: $F(1, 29) = 1.301$, $p = .263$). We consider this a spurious rather than a meaningful interaction and do not evaluate it further. Given that there was no meaningful or consistent effect of encoding task on proportion correct across the experiments, the data for the two tasks is collapsed for all further analyses. For completeness, descriptive statistics are presented separately for each encoding task in Tables 1, 2, 4 and 5.

Experiment 1

In this experiment participants study pairs of items, one of which later served as a cue and the other as a target. The WF of the cue and the target were independently manipulated for a total of four between-list conditions: LF cue with LF target; LF cue with HF target; HF cue with LF target; and HF cue with HF target. For comparison to previous studies we allowed WF and CV to remain confounded in this Experiment, as is typical in the literature (the confound is eliminated in Experiments 3 and 4).

In the paired-associate learning paradigm, participants learn to respond with the target word in response to a specific cue word. Performance in this task is typically evaluated while participants try to master a list or after mastery has been reached, in contrast to a single study-test trial common in cued recall paradigms. The paired-associate literature consistently shows that learning is faster for HF responses, but the data are unclear with respect to cues. Some report a LF advantage and others report no difference between HF and LF cues (e.g., Hall, 1972; Modigliani & Saltz, 1969; Postman, 1962; Saltz, 1967; Underwood, 1982). Cued recall experiments, like those we conducted, evaluate memory after a single learning trial and the focus is on success or failure at retrieval. Few manuscripts evaluate the role of WF in cued recall. Those that do demonstrate a HF benefit but do not separately manipulate the WF of the cue and target (Clark & Burchett, 1994) or show null effects of both cue and target WF (Gillund & Shiffrin, 1984, Appendix B). In all the experiments just cited, WF and CV are confounded.

Contemporary models assume that low frequency words are a better self-cue, as evidenced by their superior performance in single item recognition. This may be due to the extra evidence provided by matching the uncommon features in a LF target (e.g., Shiffrin & Steyvers, 1997) or it may be due to the ease of matching the experimental context for LF words (e.g., Dennis & Humphreys, 2001; Marsh et al., 2006; Reder et al., 2000). If the search phase of cued recall operates similarly to single item recognition, then LF cues should be more effective than HF cues. Conversely, if HF words are better cues for both pre-experimental and experimental associates, as assumed in the associative hypothesis of the SAM model, then HF cues should be more effective than LF cues.

If the retrieval and generation phase of cued recall is similar to free recall, then HF targets should be easier to retrieve and generate than LF targets. LF words may be more difficult to generate, perhaps due to their unusual orthographic and phonological form (e.g., Criss & Malmberg, 2008). Evidence in favor of this hypothesis comes from naming and lexical decision studies showing that people are faster to read HF than LF words and to identify them as words (e.g., Borowsky & Besner, 1993; Duchek & Neely, 1989; Forster & Chambers, 1973; Forster & Davis, 1984; Kirsner, Milech, & Standen, 1983; Scarborough, Cortese, & Scarborough, 1977). The HF benefit for naming and lexical decision may be due to the match between the common features of HF words and perceptual noise (e.g., Schooler, Shiffrin, & Raaijmakers, 2001).

Method

Participants
Fifty-six undergraduates participated.

Materials
The word pool consisted of 800 LF and 800 HF words between 4 and 11 letters in length. HF words ranged between 9 and 13 log frequency ($M = 10.46$) and LF words ranged between 3.5 and 6 log frequency ($M = 5.22$) in the Hyperspace Analog to Language corpus (HAL; Balota et al., 2007; Lund & Burgess, 1996).

Design
The word frequency of the cue and target were independently manipulated for a total of four conditions (LF cue with LF target; LF cue with HF target; HF cue with LF target; and HF cue with HF target). The experiment consisted of four blocks, each pure with respect to condition, that is each block contained a single condition. Order of condition was randomly chosen for each participant. All

Table 4
Mean response probabilities for each condition in Experiment 3 as a function of encoding task. Word frequency of the cues and targets were orthogonally manipulated (HF = high frequency, LF = low frequency) while context variability of the pair was held constant (HCV = high context variability LCV = low context variability). Standard errors are in parentheses.

	HF target		LF target	
	HF cue	LF cue	HF cue	LF cue
Experiment 3: Sentence task				
HCV pairs				
P (correct)	0.362 (0.040)	0.422 (0.041)	0.282 (0.040)	0.307 (0.045)
P (intrusion)	0.121 (0.034)	0.101 (0.032)	0.086 (0.030)	0.130 (0.035)
P (no response)	0.516 (0.040)	0.477 (0.036)	0.632 (0.041)	0.564 (0.046)
LCV pairs				
P (correct)	0.473 (0.044)	0.428 (0.043)	0.288 (0.037)	0.241 (0.039)
P (intrusion)	0.117 (0.030)	0.126 (0.030)	0.163 (0.033)	0.198 (0.037)
P (no response)	0.409 (0.037)	0.447 (0.043)	0.549 (0.042)	0.562 (0.045)
Experiment 3: Association task				
HCV pairs				
P (correct)	0.424 (0.038)	0.474 (0.039)	0.383 (0.038)	0.365 (0.042)
P (intrusion)	0.163 (0.032)	0.135 (0.031)	0.131 (0.028)	0.135 (0.033)
P (no response)	0.413 (0.038)	0.391 (0.034)	0.485 (0.039)	0.500 (0.043)
LCV pairs				
P (correct)	0.493 (0.042)	0.435 (0.041)	0.359 (0.035)	0.319 (0.037)
P (intrusion)	0.169 (0.029)	0.126 (0.028)	0.126 (0.031)	0.154 (0.035)
P (no response)	0.339 (0.035)	0.439 (0.041)	0.515 (0.040)	0.528 (0.043)

Table 5
Mean response probabilities for each condition in Experiment 4 as a function of encoding task. Context variability of the cues and targets were orthogonally manipulated (HCV = high context variability LCV = low context variability) while word frequency of the pair was held constant (HF = high frequency, LF = low frequency). Standard errors are in parentheses.

	HCV target		LCV target	
	HCV cue	LCV cue	HCV cue	LCV cue
Experiment 4: Sentence task				
HF pairs				
P (correct)	0.348 (0.032)	0.480 (0.036)	0.348 (0.035)	0.510 (0.037)
P (intrusion)	0.133 (0.023)	0.113 (0.020)	0.084 (0.013)	0.126 (0.018)
P (no response)	0.519 (0.035)	0.407 (0.034)	0.567 (0.035)	0.364 (0.031)
LF pairs				
P (correct)	0.286 (0.038)	0.325 (0.040)	0.320 (0.037)	0.343 (0.038)
P (intrusion)	0.131 (0.021)	0.123 (0.021)	0.099 (0.018)	0.163 (0.023)
P (no response)	0.582 (0.039)	0.552 (0.041)	0.581 (0.037)	0.493 (0.038)
Experiment 4: Association task				
HF pairs				
P (correct)	0.404 (0.034)	0.559 (0.037)	0.404 (0.037)	0.480 (0.039)
P (intrusion)	0.102 (0.024)	0.094 (0.021)	0.081 (0.014)	0.104 (0.019)
P (no response)	0.494 (0.037)	0.346 (0.035)	0.515 (0.037)	0.417 (0.032)
LF pairs				
P (correct)	0.324 (0.040)	0.359 (0.042)	0.330 (0.039)	0.359 (0.040)
P (intrusion)	0.122 (0.022)	0.106 (0.022)	0.050 (0.018)	0.059 (0.024)
P (no response)	0.554 (0.041)	0.535 (0.043)	0.620 (0.039)	0.581 (0.040)

remaining details were as described in the 'General method' section.

Results and discussion

A 2 × 2 repeated measures ANOVA (WF of the cue × WF of the target) was performed on the proportion of correct responses. As shown in Fig. 1, HF targets were better recalled than LF targets, $F(1, 55) = 93.97$, $p < .001$, consistent with our hypothesis. There was also a small effect of cue WF where HF cues were more effective than LF cues: $F(1, 55) = 11.83$, $p = .001$. The interaction of cue and target frequency was not significant: $F(1, 55) = 0.068$, $p = .796$. The benefit for HF cues is consistent with the associative hypothesis of SAM but inconsistent with the single item recognition data.

The benefit for HF targets and HF cues could be accompanied by a failure to generate any response or to generating an incorrect response when the target or cue is a LF word. Repeated measures ANOVAs on both types of errors were conducted and the data are reported in Table 1. There were no main effects or interactions for the percent of

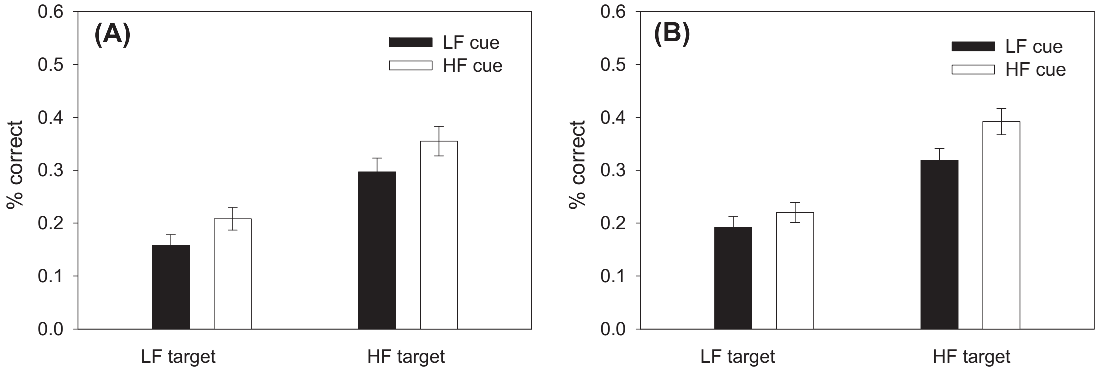

Fig. 1. Probability of correctly recalling the target item in Experiments 1 (panel A) and 2 (panel B). The word frequency of the cues and targets were orthogonally manipulated (LF = low frequency; HF = high frequency). The study lists in Experiment 1 were pure with respect to word frequency condition and the lists in Experiment 2 were mixed. The data are collapsed across encoding task within each experiment. Error bars represent one standard error above and one below the mean.

intrusions as a function of cue or target WF (all F's < 3.42 and all p's > .07). Instead, LF targets seem to fare worse than HF targets because participants are more likely to give no response when the target word is LF, $F(1, 55) = 81.00$, $p < .001$ and when the cue is LF, $F(1, 55) = 12.34$, $p = .001$. There was no interaction $F(1, 55) = 1.68$, $p = .201$.

Experiment 2

In Experiment 1 we used pure lists and showed a large advantage for HF targets and a small advantage for HF cues. This is consistent with an overall benefit for HF words in free recall of pure lists typically found in the literature. However, the two lists that were necessarily mixed (HF cue, LF target and vice versa) also showed a benefit for HF targets. In this experiment, we seek to replicate our findings from Experiment 1 using mixed lists containing all four types of pairs.

Method

Participants

Fifty-six people participated.

Materials

The word pool was the same as Experiment 1.

Design

The design was identical to that described in Experiment 1 with one exception. The study lists were mixed with respect to condition; each study–test block contained five trials of each condition type. Trial order was randomly intermixed for each participant.

Results and discussion

In the current experiment, we found the same pattern of results with mixed lists as we observed in Experiment 1 using pure lists (Fig. 1; Table 2). High frequency targets were better recalled than LF targets, $F(1, 58) = 116.66$, $p < .001$ and HF cues were more effective than LF cues, $F(1, 58) = 13.30$, $p = .001$. The interaction of cue and target frequency was not significant: $F(1, 58) = 2.66$, $p = .109$.

As in Experiment 1, the poorer performance for LF targets is accompanied by more failures to provide an answer, rather than providing incorrect answers, $F(1, 58) = 71.16$, $p < .001$, with a marginal main effect of cue WF, $F(1, 58) = 3.90$, $p = .053$, and no interaction, $F(1, 58) = 3.09$, $p = .084$. For intrusions, all F's < 3.33 and all p's > .073.

Experiment 3

Normative word frequency and context variability are highly correlated in the natural environment and likely in the word pools used in Experiments 1 and 2 (Adelman, Brown, & Quesada, 2006; Steyvers & Malmberg, 2003). We cannot be sure that the benefit for HF targets is due to the actual frequency of the words or CV (or some other property correlated with WF, see Nelson & McEvoy, 2000). Both frequency and CV carry heavy theoretical importance, especially in models of single item recognition. In this experiment, the WF of the cues and targets are manipulated as before, but CV is held constant within pairs of items. The question of interest is whether the advantage of HF cues and targets observed in our earlier experiments will persist when CV is held constant.

Method

Participants

Fifty-seven people participated.

Materials

The Steyvers and Malmberg (2003) stimuli were used. They created four sets of stimuli where CV and WF were orthogonally varied: HCV and HF, HCV and LF, LCV and HF, and LCV and LF. The WF word sets were approximately equal in CV and the CV word sets were approximately equal in WF. The CV and WF counts came from the Touchstone Applied Science Associates (TASA) corpus (Landauer, Foltz, & Laham, 1998). We were cognizant of the potential for selection effects in this situation, however, two facts somewhat allay these concerns. First, these same stimuli have been used in free recall and single item recognition studies (described in the 'Introduction' section) that showed separate effects of CV and WF. Second, we computed frequency of these words from the HAL database (used in Experiments 1 and 2) and the Google database. As shown in Table 3, the values between HF and LF categories are large and the differences between HCV and LCV categories are small or absent for all corpora.[3]

Design

The word frequency of the cues and targets were orthogonally manipulated as in Experiments 1 and 2, but now CV was held constant. For a given WF condition, all words were drawn from the LCV or the HCV set. There were four cue–target combinations: LF cue with LF target, LF cue with HF target, HF cue with LF target, and HF cue with HF target, which could consist of either all LCV or all HCV words producing a total of eight conditions. All study lists were pure with respect to condition, that is, each block contained a single condition. Participants began with the four LCV blocks or the four HCV blocks, order counterbalanced across participants. Within the CV blocks, the four conditions were randomly ordered for each participant. The study and test lists consisted of 18 trials each. Encoding time was 2.75 s per pair, slightly faster than the previous experiments to keep the duration of the experiment under an hour. All remaining details are described in the 'General method' section.

Results and discussion

For each response outcome (correct, incorrect, no response) a 2 × 2 × 2 (CV of pair, WF of cue, and WF of target) ANOVA was conducted. As shown in Fig. 2 and Table 4, the benefit for HF targets was replicated, when CV was held constant. More correct responses were given for HF than LF targets: $F(1, 56) = 88.90$, $p < .001$. The HF cueing benefit observed in Experiments 1 and 2 was eliminated: $F(1, 56) = 0.632$, $p = .430$. Note that the difference between values for HF and LF categories in Experiments 3 and 4 is substantially smaller than the difference in Experiments 1 and 2 (see Table 3). Nevertheless the HF target benefit was replicated indicating this is a robust effect. The elimination of the HF cue benefit may be due to holding CV constant or to the smaller difference between WF categories. We cannot discriminate between these two possibilities. There was no main effect of CV ($F(1, 56) = 0.013$, $p = .910$), but it interacted with WF of the cue ($F(1, 56) = 10.56$, $p = .002$) and WF of the target ($F(1, 56) = 8.29$, $p = .006$) such that any difference in percent correct between LF and HF words was magnified for LCV relative to HCV pairs, as displayed in Fig. 2. No other interactions

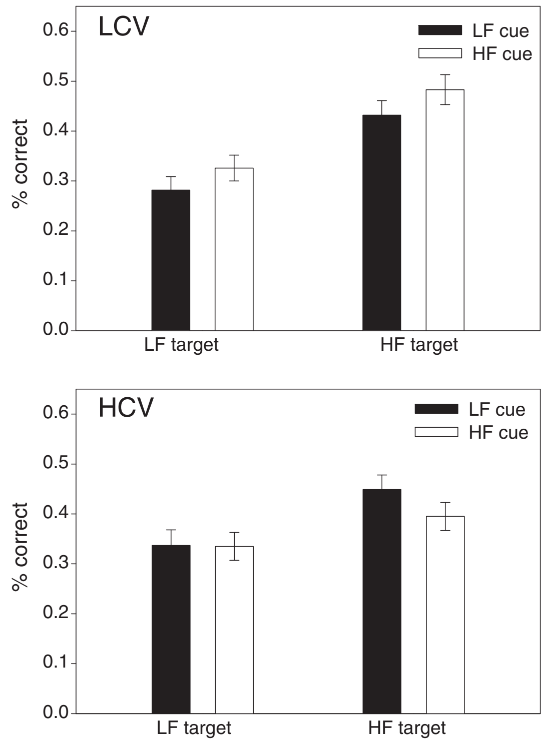

Fig. 2. Probability of correctly recalling the target item in Experiment 3. The word frequency of the cues and targets were orthogonally manipulated (LF = low frequency; HF = high frequency). Context variability was held constant for all items in a pair at either a low or high value (LCV = low context variability; HCV = high context variability). The data are collapsed across encoding task. Error bars represent one standard error above and one below the mean.

were significant (all *F*s < 1.699 and *p*'s > .198) for correct responses.

Replicating the earlier experiments, a higher rate of correct recall for HF targets was accompanied by a lower rate of trials on which no response was given: $F(1, 56) = 106.59$, $p < .001$. The CV by cue WF interaction in correct responses was accompanied by an interaction for trials with no response, $F(1, 56) = 11.02$, $p = .002$. There were no other main effects or interactions for the proportion of trials with no response (all *F*s < 2.72 and *p*'s > .104).

The analysis of intrusions revealed a very small effect of CV such that slightly more intrusions were made during low CV than high CV lists, $F(1, 56) = 5.37$, $p = .024$, an interaction between WF of the cue and target, $F(1, 56) = 9.78$, $p = .003$, and a marginal interaction between CV and WF of the target, $F(1, 56) = 3.70$, $p = .060$. There were no other main effects or interactions (all *F*'s < 0.612 and *p*'s > .437) for intrusions.

Experiment 4

All experiments reported in this manuscript demonstrate a benefit for HF targets. This is consistent with the hypothesis that HF words are more accessible, perhaps because their lexical features are more common (e.g., Landauer & Streeter, 1973) and more likely to be generated

[3] The data from Experiments 3 and 4 were also informally evaluated using path analysis to investigate the role of target and cue WF, target and cue CV, and the interaction of WF and CV on accuracy. In the analysis we used individual WF and CV values rather than treating WF and CV as categorical variables. Through this exercise we confirmed the ANOVA findings, namely that WF of targets and CV of cues are the primary factors influencing successful recall. Neither interaction term was a significant contributor to performance. Details of the analysis are available by request.

(e.g., Anderson & Bower, 1972). Surprisingly and in contrast to predictions, the WF of the cue had a much smaller impact which was eliminated when CV was held constant. In all models considered here, properties of the cue are critical to performance. SAM assumes that HF cues are better able to retrieve other items (Gillund & Shiffrin, 1984). In REM, LF cues are a better match to the memory trace containing the cue because their uncommon features provide more evidence in favor of a match than do common features found in HF cues (e.g., Malmberg et al., 2002). SAC and BCDMEM both attribute at least part of the effects of WF to context variability (Dennis & Humphreys, 2001; Reder et al., 2000). In this experiment, we evaluate the impact of CV on cues and targets when WF is controlled.

Despite a correlation between WF and CV, the two measures independently contribute to memory performance across a wide range of tasks including single item recognition, free recall, source memory, naming and lexical decision (Adelman et al., 2006; Cook et al., 2006; Hicks et al., 2005; Marsh et al., 2006; Steyvers & Malmberg, 2003). This holds when concreteness, imageability, ambiguity, and WF are taken into account in the analyses (Adelman et al., 2006) or held constant (Marsh et al., 2006). In some tasks, WF and CV have similar patterns of data such as higher HR and lower FARs in single item recognition and superior memory for the source for LF and LCV words (Cook et al., 2006). However, in other tasks the two variables do not act in concert. For example, CV predicts lexical decision and naming response times. Accounting for CV in these tasks eliminates any effect of WF that was present when WF and CV were confounded (Adelman et al., 2006). In free recall, performance is better for LCV than HCV but better for HF than LF words (Hicks et al., 2005) and this is replicated regardless of whether or not concreteness is held constant (Marsh et al., 2006).

To the best of our knowledge no experiments have manipulated CV in a cued recall task. However, the data from Marsh et al. (2006) are particularly relevant. They conducted experiments where the context at study and test matched or mismatched and showed that the LCV advantage in free recall is eliminated in the mismatched condition. They reasoned that the advantage for LCV words across episodic tasks was due to a stronger association between the LCV items and the list context. Therefore, if the item-to-context storage was interrupted, LCV items should not be better recalled. To evaluate this hypothesis, Marsh et al. (2006) adopted an encoding task where participants had to attend to the relationship between items by stating a similarity between the current study item and the study word presented in the previous trial. They intended this task to encourage storage of item-to-item associations and prevent participants from forming item-to-context associations. The results supported their predictions: the LCV advantage was eliminated for both the matched and mismatched conditions.

In the cued recall experiments presented here, an encoding task was chosen to encourage participants to form inter-item associations and indeed the memory test requires participants to remember the co-occurrence of pairs of items. Thus the Marsh et al. (2006) prediction is no effect of the CV of the cue or target in a cued recall paradigm. In other words, because the item-to-item associations are critical to performance in cued recall and are encouraged during encoding, LCV words should not benefit from their superior item-to-context associations and a null effect of CV is predicted.

BCDMEM is intended for single item recognition of words, which is assumed to be a context-noise task (cf., Criss & Shiffrin, 2004a). Strictly speaking, we cannot generate predictions for cued recall from BCDMEM. Remember however, we assume that cued recall is a multi-step process with cueing followed by sampling and retrieval. We further assume that the cueing process proceeds like single item recognition. Under that reasonable assumption, BCDMEM would predict a benefit for LCV over HCV cues. We cannot speculate about BCDMEM predictions for the retrieval process because the model does not apply to free recall.

Finally, consider SAC (Reder et al., 2000). There are three types of nodes in SAC: the concept node represents the pre-experimental familiarity of an item and the level of familiarity is derived from word frequency; context nodes represent contexts in which items have occurred including one for the experiment itself; the episode node is the binding between the context and concept nodes. The episode node is where recollection and recall take place (i.e., remembering the concept in the experimental context; Reder et al., 2000, 2007) and the activation of that node is partially determined by the total number of contexts bound to the concept. More activation spreads to episode nodes from concepts with a smaller number of prior contexts. SAC, then, seems to predict a benefit for LF/LCV words both as cues and as targets.

The design of this experiment is similar to Experiment 3 with the role of WF and CV reversed. The CV of the cue and the target were independently manipulated and WF was held constant at two levels: HF and LF.

Method

Participants

Sixty-three people participated.

Materials

The same word pool as Experiment 3 was used.

Design

Context variability of the cues and targets was orthogonally manipulated and WF was held constant across CV conditions. Similar to Experiment 3 there were four cue–target combinations: LCV cue with LCV target, LCV cue with HCV target, HCV cue with LCV target, and HCV cue with HCV target, which could consist of either all HF or all LF words producing a total of eight conditions. Blocks were pure with respect to condition. Participants began with the four LF blocks or the four HF blocks, order counterbalanced. Within each blocks, the four conditions and the words within each condition were randomly ordered anew for each participant. All remaining details are described in the 'General method' section.

Results and discussion

For each response outcome (correct, incorrect, no response) a $2 \times 2 \times 2$ (WF of pair, CV of cue, and CV of target) repeated measures ANOVA was conducted. Mean values are reported in Table 5. Consistent with earlier experiments, performance was better for HF than LF pairs, $F(1, 62) = 49.076$, $p < .001$. This was accompanied by fewer trials where no response was given to HF compared to LF pairs, $F(1, 62) = 35.55$, $p < .001$.

As shown in Fig. 3, the results are consistent with the predictions of BCDMEM, but not Marsh et al. (2006). LCV words are better cues than HCV words: $F(1, 62) = 64.75$, $p < .001$. The benefit for LCV cues is exaggerated for HF compared to LF pairs, as evidenced by an interaction between WF and cue CV: $F(1, 62) = 17.18$, $p < .001$. Both of these findings are accompanied by a reduction in trials where no response was provided for the main effect of cue CV ($F(1, 62) = 49.64$, $p < .001$) and for the interaction between WF of the pair and cue CV ($F(1, 62) = 15.46$, $p < .001$.) There were slightly more intrusions to HCV targets ($M = .116$) than LCV targets ($M = .097$), $F(1, 62) = 8.01$, $p = .006$ and an interaction between CV of the cue and target on intrusions, $F(1, 62) = 12.86$, $p = .001$) such that intrusions are slightly higher when the CV of the cue and target match. There were no other main effects or interactions (all F's < 2.7137 and p's > .149) for any of the three response types. Of note, the CV of the target had no effect on accuracy.

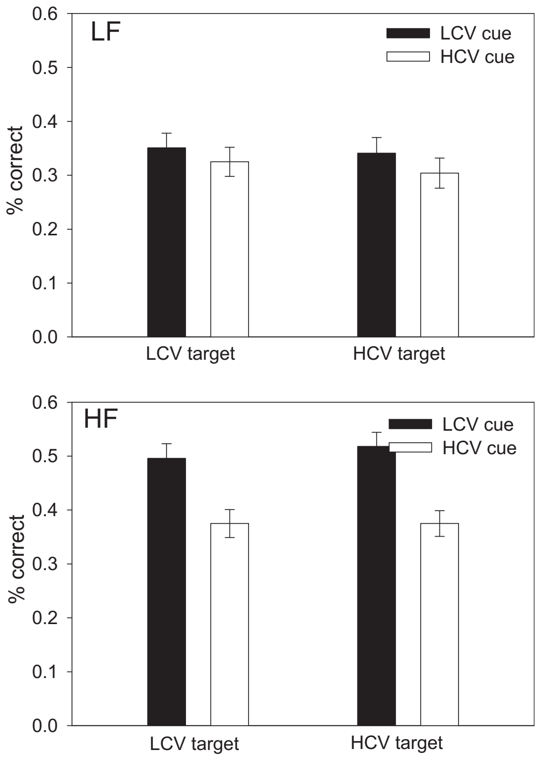

Fig. 3. Probability of correctly recalling the target item in Experiment 4. The context variability of the cues and targets were orthogonally manipulated (LCV = low context variability; HCV = high context variability). Word frequency was held constant for all items in a pair at either a low or high value (LF = low frequency; HF = high frequency). The data are collapsed across encoding task. Error bars represent one standard error above and one below the mean.

Both Experiments 3 and 4 contain interactions such that the effect of WF is larger for LCV than HCV words (e.g., the combination of HF and LCV boosts memory). This is a common finding when these variables are manipulated together (e.g., Cook et al., 2006; Hicks et al., 2005; Marsh et al., 2006; Steyvers & Malmberg, 2003). The memory signal is multiplicative, so it is not too surprising that the effect of LCV words (which are superior in recognition) and HF words (which are superior in recall) combine to improve memory more than either variable alone. It is possible that this is a vonRestorff-like finding (e.g., Erickson, 1965). By definition, HF-LCV words occur frequently and in the same type of context most of the time. The experimental context is perhaps a particularly unusual circumstance in which to see HF-LCV words, resulting in improved encoding and higher accuracy in a memory task. Of course this is all speculation. More importantly, the interactions reported here are uninterpretable and should be considered with great caution (Wagenmakers, Kryptos, Criss, & Iverson, submitted for publication). As Loftus (1978) pointed out, interactions like these cannot be interpreted because they depend on the measurement scale and can be removed with transformations (whether those transformations are in the data space, in the brain, etc.). It is possible that such interactions are due to a true underlying mechanism of the cognitive system. It is also possible that such interactions are not theoretically interesting but rather a byproduct of the scale used to produce or measure performance. Further research is needed to establish whether the interaction between WF and CV is theoretically meaningful.

General discussion

These experiments demonstrate that both WF and CV contribute in a meaningful way to cued recall performance: high WF targets are more accessible and more likely to be correctly recalled and low CV cues are more likely to prompt recall of their study partner. In three experiments, the WF of the cue and target were orthogonally manipulated in mixed and pure lists, when CV was controlled and when it was not. Performance was better for HF compared to LF targets. A small benefit for HF cues was eliminated when CV was controlled. The fourth experiment showed a LCV cueing advantage and null effect of target CV when WF was held constant. These data clarify earlier experiments where the effect of WF on cues and targets in cued recall was ambiguous (e.g., Clark & Burchett, 1994; Gillund & Shiffrin, 1984).

Theoretical implications

The broad purpose of this research is to begin to bridge the gap between models of single item recognition and models of free recall. Toward that goal we adopted a model of cued recall that reflects both REM (Shiffrin & Steyvers,

1997) a model primarily applied to single item recognition and SAM a model most successful at accounting for free recall (Raaijmakers & Shiffrin, 1981). We assumed cued recall involves comparing a cue to memory, sampling a single memory trace (which contains item information for the cue and target, context information, and possibly associative information, cf. Criss & Shiffrin, 2005), and attempting to retrieve the target item in that trace. We further assumed that cueing is similar to single item recognition and retrieval is similar to free recall, which allowed us to generate predictions based on contemporary models for single item recognition, theories of free recall, and the SAM model.

The lack of a LF cue advantage is inconsistent with predictions of most models of single item recognition and this would seem to be a particularly difficult problem to resolve when building a multi-task model (Dennis & Humphreys, 2001; McClelland & Chappell, 1998; Shiffrin & Steyvers, 1997; Reder et al., 2000). One possible explanation follows. LF words demand more time, effort, and/or attention to encode and identify their relatively unusual orthography (e.g., Criss & Malmberg, 2008; Landauer & Streeter, 1973; Malmberg et al., 2002; Malmberg & Nelson, 2003; McDaniel & Bugg, 2008). This idea, sometimes called the elevated attention hypothesis, has been used to explain the mixed list paradox (DeLosh & McDaniel, 1996; May & Sande, 1982; Merritt et al., 2006; Watkins et al., 2000) and the elimination of the LF HR advantage in single item recognition when participants are given a task that directs attention to semantic, associative, or other non-orthographic features (Criss & Shiffrin, 2004c; Criss & Malmberg, 2008; Glanc & Greene, 2007). In the studies reported here, participants focus on binding together the pair items during encoding. Such a task may eliminate or reduce the LF cueing advantage that would otherwise be present (e.g., Malmberg & Nelson, 2003). Eliminating an associative task at encoding and/or not disclosing the nature of the test may reduce overall performance in cued recall but preserve a LF cueing advantage.

Recognition models that emphasize the importance of pre-experimental contexts (like BCDMEM and SAC) are consistent with our finding of a LCV cue advantage. While neither model explicitly equates the number of prior contexts with CV (e.g., operationalized as the number of *different* prior contexts), it seems a natural extension of these models.

The presence of a HF target advantage is problematic for several models. The hypothesis that properties of common words make them easier to generate or are more readily available has been around for some time (e.g., Underwood & Schulz, 1960). Findings that intrusions are more likely to be HF words provide converging evidence for the accessibility hypothesis (e.g., Madan, Glaholt, & Caplan, 2010). Despite evidence for faster access to HF words from naming, perceptual identification, and lexical decision tasks, the mechanism for an availability advantage for HF words has not been implemented in contemporary models of episodic memory (Borowsky & Besner, 1993; Duchek & Neely, 1989; Forster & Chambers, 1973; Forster & Davis, 1984; Kirsner et al., 1983; Scarborough et al., 1977). Incorporating the accessibility hypothesis into formal models of memory is long overdue and seems a necessary step in building a comprehensive model of episodic memory.

Next consider implications for theories of free recall. According to the order-encoding hypothesis, there exists a tradeoff between item and order information. LF items attract more attention to encoding item features and consequently less attention is available to encode order information, relative to HF words (DeLosh & McDaniel, 1996; Merritt et al., 2006; Toglia & Kimble, 1976). Predictions of the order-encoding hypothesis are ambiguous with respect to the cue. The cue consists of item information, which should benefit LF words. To the extent that order information provided by the cue helps guide participants to the correct memory trace (i.e., both the cue and the target occurred in the same order with respect to the other encoded pairs), HF words should make better cues. For targets the order-encoding hypothesis predicts a LF advantage, under the assumption that successful retrieval is determined by item information.

The recency hypothesis states that HF words are better recalled because they are rehearsed more often and closer to the time of test than LF words (Tan & Ward, 2000; Ward et al., 2003). To the extent that HF words are rehearsed individually in a cued recall task, this hypothesis predicts better performance for both cues and targets that are HF. However, given that the task is to recall an item when cued with it's partner, participants might rehearse pairs of items rather than individual items (e.g., Atkinson & Shiffrin, 1968; Raaijmakers & Shiffrin, 1981). If the recency hypothesis extends to pairs, then accuracy should be higher for HF than LF or mixed pairs. The current data do not support either prediction.[4] The recency hypothesis also incorporates the associative hypothesis of SAM that HF words have stronger inter-item associations discussed next.

SAM assumes that HF words are better associated to other items and therefore better cues. The model assumes no WF difference in the sampling and recovery (Gillund & Shiffrin, 1984). Therefore, HF words are predicted to be better recalled than LF words due to the HF cueing advantage in SAM, not an advantage in the ability to retrieve target words (Gillund & Shiffrin, 1984). The first two experiments did show a small HF cueing advantage but this was not replicated when CV is controlled. The benefit for HF targets, regardless of cue type, suggests recovery is influenced by WF, contrary to SAM.

Generate-recognize models are another possible process underlying cued recall performance (e.g., Anderson & Bower, 1972; Jacoby & Hollingshead, 1990). In a generate-recognize model, possible target words are generated and are then subjected to a recognition check. If the generated word is recognized as having been on the list, it is reported. One primary difference between the SAM/REM-inspired model of cued recall that we describe in this manuscript and a generate-recognize model is the recognition check. How might the recognition check alter predictions for WF and CV of the cue and target? LF and LCV targets are better recognized (e.g., Steyvers & Malmberg, 2003) and thus they should benefit from the secondary recognition check. Even if more HF words are generated

[4] Observation of serial positive curves indicate no consistent effect of serial position on cued recall performance.

(possibly due to heightened accessibility), they would be less likely to be recognized than LF or LCV targets. Thus adding a recognition check to the cued recall model does not seem to solve the problems presented by the current set of data.

Associative information

One factor neglected in the models described above is the strategic generation of emergent association features that include information beyond that contained by the individual items (e.g., Criss & Shiffrin, 2004b, 2005; convolutions in Murdock, 1997; higher order units in SAM, Murnane & Shiffrin, 1991, class attributes in Underwood, 1969). Such information is present in a pair stimulus and likely contributes to an associative recognition task where memory for pairs of items is tested. In cued recall, use of emergent associative information would require participants to generate such information when presented with a single item cue and then use it to help accept or reject words retrieved from memory before overt generation of the target. WF effects in associative recognition are ambiguous, with some findings of an HF benefit (e.g., Clark & Burchett, 1994) and other reports of no difference between HF and LF pairs (e.g., Hockley, 1994). At this time, there is insufficient evidence regarding the use of emergent associations in cued recall and insufficient evidence regarding the role of WF in such associations.

Madan et al. (2010) asked whether associations are symmetric or directional (e.g., is using A to retrieve B similar to using B to retrieve A; Kahana, 2002) and whether associative symmetry is modulated by item properties. Symmetric associations are consistent with models assuming emergent associative features (e.g., Criss & Shiffrin, 2004b, 2005; Murdock, 1997; Murnane & Shiffrin, 1991). WF was one item property they evaluated and in Experiment 1a, they separately manipulated the WF of cues and of targets in cued recall. Their results support symmetric (or holistic) associations and further show that associative symmetry is not disrupted by WF of the items. The details of their procedure differed from ours in several ways, most significantly they tested memory for each pair twice in two consecutive tests (i.e., a different member of the pair served as the cue in each test) and the study lists consisted of just eight pairs. Despite methodological differences, their data are consistent with ours: a benefit for HF targets and a null effect of cue WF. They report a statistical interaction between target WF and pair type (mixed vs. pure frequency) but their subsequent modeling showed that the results are determined almost entirely by the WF of the target (i.e., pages 54 and 58 of their manuscript). A survey of the literature including the current finding of no interaction and the Madan et al. finding indicate that the jury is still out on whether there exists an advantage in forming associations between HF words. This is an important topic that deserves the attention of future research.

Summary

The four experiments reported here demonstrate a benefit for HF over LF targets, perhaps due to the high availability of common features found in HF targets. Contrary to most model predictions, LF cues are not consistently better than HF cues. Consistent with broad principles of some models, LCV words are better cues than HCV words. The CV of targets has no effect on cued recall performance. The data make it clear that developing a combined model of recall and recognition is not as simple as taking the cueing process from the latter and combining it with the retrieval process in the former. Rather, the data place strong constraints on the form of a combined model. Such a model requires independent contributions from CV and WF, cueing processes that take into account both variables, and a generating process that is modulated by WF.

Acknowledgment

We thank the PRIDE program in the Department of Psychology at Syracuse University for support. We thank Dr. Aesoon Park for oversight and guidance with the path analysis. This research was funded by the National Science Foundation under Grant #0951612.

References

Adelman, J., Brown, G., & Quesada, J. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science, 17*, 814–823.
Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review, 79*, 97–123.
Atkinson, R., & Shiffrin, R. (1968). Human memory: A proposed system and its control processes. *The psychology of learning and motivation: II*. Oxford, England: Academic Press.
Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., et al. (2007). The English lexicon project. *Behavior Research Methods, 39*, 445–459.
Borowsky, R., & Besner, D. (1993). Visual word recognition: A multistage activation model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 813–840.
Bradley, M., Greenwald, M., Petry, M., & Lang, P. (1992). Remembering pictures: Pleasure and arousal in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 379–390.
Clark, S., & Burchett, R. (1994). Word frequency and list composition effects in associative recognition and recall. *Memory & Cognition, 22*, 55–62.
Cleary, A., Morris, A., & Langley, M. (2007). Recognition memory for novel stimuli: The structural regularity hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 379–393.
Cook, G., Marsh, R., & Hicks, J. (2006). The role of recollection and familiarity in the context variability mirror effect. *Memory & Cognition, 34*, 240–250.
Criss, A., & Malmberg, K. (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., & Shiffrin, R. (2004a). Context noise and item noise jointly determine recognition memory: A comment on Dennis and Humphreys (2001). *Psychological Review, 111*, 800–807.
Criss, A., & Shiffrin, R. (2004b). Pairs do not suffer interference from other types of pairs or single items in associative recognition. *Memory & Cognition, 32*, 1284–1297.
Criss, A., & Shiffrin, R. (2004c). 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.
Criss, A., & Shiffrin, R. (2005). List discrimination in associative recognition and implications for representation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 1199–1212.
DeLosh, E., & McDaniel, M. (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. (2001). A context noise model of episodic word recognition. *Psychological Review, 108*, 452–478.
Diller, D., Nobel, P., & Shiffrin, R. (2001). An ARC–REM model for accuracy and response time in recognition and recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 414–435.
Duchek, J., & Neely, J. (1989). A dissociative levels-of-processing interaction in episodic × word-frequency recognition and lexical decision tasks. *Memory & Cognition, 17*, 148–162.
Erickson, R. L. (1965). Differential effects of stimulus and response isolation in paired-associate learning. *Journal of Experimental Psychology, 69*, 317–323.
Forster, K., & Chambers, S. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior, 12*, 627–635.
Forster, K., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 680–698.
Freeman, E., Heathcote, A., Chalmers, K., & Hockley, W. (2010). Item effects in recognition memory for words. *Journal of Memory and Language, 62*, 1–18.
Gillund, G., & Shiffrin, R. (1981). Free recall of complex pictures and abstract words. *Journal of Verbal Learning and Verbal Behavior, 20*, 575–592.
Gillund, G., & Shiffrin, R. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91*, 1–67.
Glanc, G. A., & Greene, R. L. (2007). Orthographic neighborhood size effects in recognition memory. *Memory & Cognition, 35*, 365–371.
Glanzer, M., & Adams, J. (1985). The mirror effect in recognition memory. *Memory & Cognition, 13*, 8–20.
Glanzer, M., & Adams, J. (1990). The mirror effect in recognition memory: Data and theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 5–16.
Greene, R. (2004). Recognition memory for pseudowords. *Journal of Memory and Language, 50*, 259–267.
Gregg, V. (1976). Word frequency, recognition and recall. In J. Brown (Ed.), *Recall and recognition* (pp. 183–216). Oxford, England: John Wiley & Sons.
Gregg, V., Montgomery, D., & Castaño, D. (1980). Recall of common and uncommon words from pure and mixed lists. *Journal of Verbal Learning and Verbal Behavior, 19*, 240–245.
Grider, R., & Malmberg, K. (2008). Discriminating between changes in bias and changes in accuracy for recognition memory of emotional stimuli. *Memory & Cognition, 36*, 933–946.
Hall, J. F. (1954). Learning as a function of word frequency. *American Journal of Psychology, 67*, 138–140.
Hall, J. (1972). Associative strength and word frequency as related to stages of paired-associate learning. *Canadian Journal of Psychology, 26*, 252–258.
Hicks, J., Marsh, R., & Cook, G. (2005). An observation on the role of context variability in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 1160–1164.
Hockley, W. (1994). Reflections of the mirror effect for item and associative recognition. *Memory & Cognition, 22*, 713–722.
Hulme, C., Stuart, G., Brown, G. D. A., & Morin, C. (2003). High- and low-frequency words are recalled equally well in alternating lists: Evidence for associative effects in serial recall. *Journal of Memory & Language, 49*, 500–518.
Jacoby, L. L., & Hollingshead, A. (1990). Toward a generate/recognize model of performance on direct and indirect tests of memory. *Journal of Memory & Language, 29*, 443–454.
Kahana, M. (2002). Associative symmetry and memory theory. *Memory & Cognition, 30*, 823–840.
Kapucu, A., Rotello, C., Ready, R., & Seidl, K. (2008). Response bias in 'remembering' emotional stimuli: A new perspective on age differences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 703–711.
Kirsner, K., Milech, D., & Standen, P. (1983). Common and modality-specific processes in the mental lexicon. *Memory & Cognition, 11*, 621–630.
Landauer, T., Foltz, P., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes, 25*, 259–284.
Landauer, T., & Streeter, L. (1973). Structural differences between common and rare words: Failure of equivalence assumptions for theories of word recognition. *Journal of Verbal Learning and Verbal Behavior, 12*, 119–131.
Loftus, G. (1978). On interpretation of interactions. *Memory & Cognition, 6*, 312–319.
Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments & Computers, 28*, 203–208 (R).
Madan, C., Glaholt, M., & Caplan, J. (2010). The influence of item properties on association-memory. *Journal of Memory and Language, 63*, 46–63.
Malmberg, K., & Nelson, T. (2003). The word frequency effect for recognition memory and the elevated-attention hypothesis. *Memory & Cognition, 31*, 35–43.
Malmberg, K., Steyvers, M., Stephens, J., & Shiffrin, R. (2002). Feature frequency effects in recognition memory. *Memory & Cognition, 30*, 607–613.
Marsh, R., Meeks, J., Hicks, J., Cook, G., & Clark-Foos, A. (2006). 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.
May, R., Cuddy, L., & Norton, J. (1979). Temporal contrast and the word frequency effect. *Canadian Journal of Psychology, 33*, 141–147.
May, R., & Sande, G. (1982). Encoding expectancies and word frequency in recall and recognition. *American Journal of Psychology, 95*, 485–495.
McClelland, J., & Chappell, M. (1998). Familiarity breeds differentiation: A subjective-likelihood approach to the effects of experience in recognition memory. *Psychological Review, 105*, 724–760.
McDaniel, M., & Bugg, J. (2008). Instability in memory phenomena: A common puzzle and a unifying explanation. *Psychonomic Bulletin & Review, 15*, 237–255.
Merritt, P., DeLosh, E., & McDaniel, M. (2006). Effects of word frequency on individual-item and serial order retention: Tests of the order-encoding view. *Memory & Cognition, 34*, 1615–1627.
Modigliani, V., & Saltz, E. (1969). Evaluation of a model relating Thorndike–Lorge frequency and m to learning. *Journal of Experimental Psychology, 82*, 584–586.
Murdock, B. (1997). Context and mediators in a theory of distributed associative memory (TODAM2). *Psychological Review, 104*, 839–862.
Murnane, K., & Shiffrin, R. (1991). Word repetitions in sentence recognition. *Memory & Cognition, 19*, 119–130.
Nelson, D., & McEvoy, C. (2000). What is this thing called frequency? *Memory & Cognition, 28*, 509–522.
Nelson, D., Reed, V., & McEvoy, C. (1977). Learning to order pictures and words: A model of sensory and semantic encoding. *Journal of Experimental Psychology: Human Learning and Memory, 3*, 485–497.
Onyper, S., Zhang, Y., & Howard, M. W. (2010). Some-or-none recollection: Evidence from item and source memory. *Journal of Experimental Psychology: General, 139*, 341–364.
Ozubko, J., & Joordens, S. (2007). The mixed truth about frequency effects on free recall: Effects of study list composition. *Psychonomic Bulletin & Review, 14*, 871–876.
Paivio, A. (1971). *Imagery and verbal processes*. Oxford, England: Holt, Rinehart & Winston.
Postman, L. (1962). The effects of language habits on the acquisition and retention of verbal associations. *Journal of Experimental Psychology, 64*, 7–19.
Raaijmakers, J., & Shiffrin, R. (1981). Search of associative memory. *Psychological Review, 88*, 93–134.
Ratcliff, R., Clark, S., & Shiffrin, R. (1990). List-strength effect: I. Data and discussion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 163–178.
Reder, L., Nhouyvanisvong, A., Schunn, C., Ayers, M., 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.
Reder, L., Oates, J., Dickinson, D., Anderson, J., Gyulai, F., Quinlan, J., et al. (2007). Retrograde facilitation under midazolam: The role of general and specific interference. *Psychonomic Bulletin & Review, 14*, 261–269.
Saltz, E. (1967). Thorndike–Lorge frequency and M of stimuli as separate factors in paired-associate learning. *Journal of Experimental Psychology, 73*, 473–478.
Scarborough, D., Cortese, C., & Scarborough, H. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception and Performance, 3*, 1–17.
Schooler, L., Shiffrin, R., & Raaijmakers, J. (2001). A Bayesian model for implicit effects in perceptual identification. *Psychological Review, 108*, 257–272.
Schulman, A. (1967). Word length and rarity in recognition memory. *Psychonomic Science, 9*, 211–212.
Shiffrin, R., Ratcliff, R., & Clark, S. (1990). List-strength effect: II. Theoretical mechanisms. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 179–195.
Shiffrin, R., & Steyvers, M. (1997). A model for recognition memory: REM – retrieving effectively from memory. *Psychonomic Bulletin & Review, 4*, 145–166.

Snodgrass, J., & McClure, P. (1975). Storage and retrieval properties of dual codes for pictures and words in recognition memory. *Journal of Experimental Psychology: Human Learning and Memory, 1*, 521–529.

Steyvers, M., & Malmberg, K. (2003). The effect of normative context variability on recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 760–766.

Stuart, G., & Hulme, C. (2000). The effects of word co-occurrence on short-term memory: Associative links in long-term memory affect short-term memory performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 796–802.

Tan, L., & Ward, G. (2000). A recency-based account of the primacy effect in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1589–1625.

Toglia, M., & Kimble, G. (1976). Recall and use of serial position information. *Journal of Experimental Psychology: Human Learning and Memory, 2*, 431–445.

Underwood, B. (1969). Attributes of memory. *Psychological Review, 76*, 559–573.

Underwood, B. (1982). Proactive interference and the simultaneous acquisition retention phenomenon. *Journal of Verbal Learning and Verbal Behavior, 21*, 142–149.

Underwood, B., & Schulz, R. (1960). *Meaningfulness and verbal learning*. Oxford, England: J.B. Lippincott.

Wagenmakers, E.-J., Kryptos, A.-M., Criss, A., & Iverson, G. (submitted for publication). On the interpretation of uninterpretable interactions: A survey of the field, 32 Years after Loftus.

Ward, G., Woodward, G., Stevens, A., & Stinson, C. (2003). Using overt rehearsals to explain word frequency effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 186–210.

Watkins, M., LeCompte, D., & 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.

Zechmeister, E. (1969). Orthographic distinctiveness. *Journal of Verbal Learning and Verbal Behavior, 8*, 754–761.