What is this thing called frequency?

DOUGLAS L. NELSON and CATHY L. McEVOY
University of South Florida, Tampa, Florida

When researchers are interested in the influence of long-term knowledge on performance, printed word frequency is typically the variable of choice. Despite this preference, we know little about what frequency norms measure. They ostensibly index how often and how recently words are experienced, but words appear in context, so frequency potentially reflects an influence of connections with other words. This paper presents the results of a large free association study as well as the results of experiments designed to evaluate the hypothesis that common words have stronger connections *to* other words. The norms indicate that common words tend to be more concrete but they do not appear to have more associates, stronger associates, or more connections among their associates. Two extralist cued recall experiments showed that, with other attributes being equal, high- and low-frequency words were equally effective as test cues. These results suggest that frequency does not achieve its effects because of stronger or greater numbers of connections *to* other words, as implied in SAM. Other results indicated that common words have more connections *from* other words, including their associates, and that free association provides a valid index of associative strength.

The purpose of this article is to address a simple question: What does printed frequency measure? Natural language frequency norms report counts of the number of times a word appears in samples of text (e.g., Kučera & Francis, 1967; Thorndike & Lorge, 1944), and words that appear more frequently tend to appear in a greater variety of contexts (Dennis, 1995). To name but a few cognitive tasks, such norms have played a significant role in research on reading, perception, lexical decision, naming, recall, and recognition for more than 50 years. Frequency norms have had a substantial influence on both practical application and theory and presumably index the accessibility of a lexical representation in memory. High-frequency words are experienced more often and, depending on the theory, such experience ostensibly strengthens a single representation for each word or increases the number of representations of that word in memory. Furthermore, because high-frequency words are experienced more often, they should on average be experienced more recently than low-frequency words. The combination of repetition and recency theoretically makes high-frequency words more retrievable for use in various cognitive tasks (Underwood & Schulz, 1960, p. 86).

These considerations imply that indices of printed frequency capture a property of the word itself and therefore measure what could be called item-specific information. This hypothesis, however, fails to recognize that words ap-

pear in text along with other words. Printed frequency may index something in addition to item-specific information, and one possibility is that it measures associative relationships among words. Given how frequency is counted, high-frequency words occur in more samples of text than low-frequency words and, as a consequence, high-frequency words may be connected with a greater variety of different words than low-frequency words. Furthermore, as illustrated in Figure 1, such connections conceivably could go in one or both of two directions. By comparison with uncommon words, common words may have more connections *to* and/or *from* other words. In the early literature the connections-*to* hypothesis was generally accepted because, in continuous association tasks, high-frequency words produced more associates (Cofer & Shevitz, 1952; Gregg, 1976, p. 186; Noble, 1952; Paivio, Yuille, & Madigan, 1968; Saltz, 1967; Winnick & Kressel, 1965). In addition, common words appear to have more connections *from* other words because a greater variety of words produce them as associates (Howes, 1957; Rubin & Friendly, 1986).

Findings reported in the early literature suggest that the directional *to*-*from* distinction may be irrelevant, but some reflection and a critical assessment of this work suggest otherwise. Reflection about the source of frequency counts suggests that there are at least two ways of thinking about what these counts might measure, what we will call the reader and writer points of view. Frequency norms are typically interpreted as reflecting the experience of an average reader. For example, Kučera and Francis (1967) selected 2,000 samples of 500 words from various sources thought to be representative of examples of words that a person might experience during reading. From the reader point of view, the directional distinction seems unnecessary because words that occur in more samples of text will presumably form more connections in both direc-

This research was supported by Grant MH16360 from the National Institute of Mental Health to D.L.N. and Grant AG13973 from the National Institute on Aging to C.L.M. Our special thanks go to Todd Leibert for his help in building the lists and collecting the data for Experiments 1 and 2, and to Nan Zhang for her help in analyzing the results of the ListChecker program. Correspondence should be addressed to D. L. Nelson, Department of Psychology, University of South Florida, Tampa, FL 33620-8200 (e-mail: nelson@luna.cas.usf.edu).

Figure 1. Connections from and to other words for high- and low-frequency words.

tions. As a result of contiguous experience, items read in more contexts should have more connections to other words and from other words. However, if we remind ourselves that each of these samples was written by an individual with a specific purpose in mind, then the directional distinction may be important. In the broadest sense, writing text is a production task guided by general plans and intended meanings (e.g., Hayes & Flower, 1986; van Dijk & Kintsch, 1983). The selection of a specific word during writing would seem to be determined primarily by its accessibility given the writer's intentions as well as words appearing earlier in the text. From the writer point of view, word frequency should be more influenced by connections *from* other words than by connections *to* other words.

The writer view of frequency norms is interesting for two reasons. First, although it assumes that the source of differential printed frequency lies in connections emanating from rather than to other words, the writer view seems compatible with the item-specific idea that frequency counts capture accessibility. Common words may be more accessible because they are experienced more often and more recently, and also because more words produce them as responses or associates. The writer view suggests that frequency counts index the accessibility of different words and that there are several different sources of such accessibility, including repetition, recency, and connections from other words. Increased access will affect the likelihood that a high-frequency word will be produced as an associate to another word, but greater access will have no effect on the likelihood that a high-frequency word will produce other words. High-frequency words are more accessible but not necessarily more productive.

A second interesting characteristic of the writer view is that it makes different predictions about free associa-

tion and cued recall than does the reader view. In the free association task, people are given stimuli such as words and are asked to write the first associated word to come to mind. Both writer and reader perspectives predict that, in free association, high-frequency words should be produced as associates *by* a greater number of different words than should low-frequency words. If high-frequency words are more accessible, then a greater variety of words should bring them to mind. These perspectives, however, differ in predictions about the effects of the frequency of the word being normed. The reader view suggests that high-frequency words will be produced by greater numbers of words and will also produce a greater variety of different words themselves when they serve as the normative stimulus. The writer perspective, in contrast, suggests that the frequency of the word being normed should have little bearing on how many different words are produced.

Early evidence favors the reader view in showing that high-frequency words tend to produce more associates. For example, an analysis of the Paivio et al. (1968) norms shows that high-frequency words (A and AA) produce an average of 6.35 associates ($n = 282$ words), whereas low-frequency words (0–5 per million) produce an average of 5.26 associates ($n = 299$). On average, high-frequency words produce an additional associate. This numerical difference is small but, more importantly, the evidence suffers from at least two shortcomings. First, continuous association was used to collect the norms, and in this procedure people produce as many associates as they can in a fixed period of time to the cue (Noble, 1952). Continuous association makes it difficult to determine whether a given word was produced because of a connection to the word being normed or because of a connection to a word produced earlier in the sequence (e.g., *cat-dog-bone*; Postman, 1970). Additionally, retrieval inhibition can build

up when multiple retrievals are based on a single cue (Roediger, 1974). The buildup of retrieval inhibition may explain why so few associates are produced in continuous as compared with discrete association procedures (Nelson & Schreiber, 1992), and it is conceivable that, in continuous association, high-frequency words are less vulnerable to inhibition because they are more accessible in memory (e.g., Rubin & Friendly, 1986).

Second, the small sample size of the earlier studies is problematic because it raises concerns about generalizability. The study with the smallest sample size evaluated only 8 words, 4 at each level of frequency (Cofer & Shevitz, 1952). Other studies evaluated 32 words (Winnick & Kressel, 1965), 77 words (Noble, 1952), 96 words (Postman, 1970), 100 words (Saltz, 1967), and 925 words (Pai-vio et al., 1968). Although the latter study examined a sizable sample of words, continuous association was used to determine how many different words were produced. These considerations suggest that there are reasons to have doubts about whether high-frequency words have connections to more words.

One purpose of the present article is to report the free association results of a much larger sample of 5,019 words (Nelson, McEvoy, & Schreiber, 1998). Participants were asked to free associate the first word to come to mind for each stimulus word. Given that only a single associate was produced, both response chaining and retrieval inhibition were eliminated and, with such a large sample, correlations between printed frequency and the number of different words produced should be more generalizable to the word population as a whole. We will refer to the index of number of different words produced as *set size*, and the question is whether common words have connections to a greater variety of different words than do uncommon words (i.e., have larger set sizes). In addition, because other indices besides set size were created, the norms will be used to address other questions about frequency. For example, do common words have connections from a greater variety of different words (see also Howes, 1957; Rubin & Friendly, 1986)? Do they tend to be more abstract? Are their associates more interconnected, more organized? Are their associates more likely to produce them as associates in turn—that is, to resonate with or to be back-connected to them? Do they produce fewer nonresponses in free association? Finally, are common words more strongly connected to related words in memory?

The last question is particularly important to the second purpose of this article. The second goal is to test experimentally the hypothesis that high-frequency words are more strongly related to words in memory than are low-frequency words. To use an example taken from Gillund and Shiffrin (1984, p. 44), the HORSE to STIRRUP connection could be stronger than the STIRRUP to HORSE connection because HORSE is a more frequently occurring word than is STIRRUP. This expectation seems counterintuitive, but they noted that our ideas about strength are biased by

free association results that would be likely to show that the opposite was true. They suggested that low-frequency words cause high-frequency words to be sampled because competing associates for such words are fewer and weaker. HORSE could have a stronger connection to STIRRUP than the reverse but produce STIRRUP less often than the reverse because HORSE is strongly connected to so many other words. In other terms, they argued that the absolute strength of the connections between high-frequency words and their associates should be stronger than the absolute strength between low-frequency words and their associates. This frequency–strength assumption plays a critical role in explaining why low-frequency words are more likely to be recognized than high-frequency words in the SAM model (Gillund & Shiffrin, 1984, p. 34). Other things being equal, the model assumes that high-frequency distractors are more strongly related to studied list words than are low-frequency distractors. As a result, high-frequency distractors seem more familiar and low-frequency words are more likely to be correctly recognized when recognition is corrected for false alarms.

Experiment 1 was designed to test the frequency–strength hypothesis directly. Participants studied a list of words presented individually for a brief period and were then given an extralist cue related to each studied word to prompt recall. Four different cues were selected for each target, representing factorial combinations of cue-to-target strength and cue frequency. Cues were strongly or weakly related to their targets according to free association data, and they were either high or low in printed frequency. The underlying rationale for this procedure assumes that this type of cued recall can provide an assessment of the strength of connection between any two words that is potentially independent of free association performance. Participants in both free association and cued recall tasks are provided with a word as a retrieval cue for another word. The essential difference is that any related word will suffice to meet the recall criterion in free association, whereas only a related word that was also in the study list will suffice in cued recall. This difference in response restriction has substantial consequences for performance, and it seems reasonable to assume that although both tasks ostensibly can be used to assess strength of connection, the two tasks may not agree as to what is a stronger cue. Given this rationale, the frequency–strength hypothesis predicts that, at each level of normative strength, high-frequency cues should be more effective than low-frequency cues. Alternatively, if free association norms adequately capture strength of connection regardless of their relative nature, then frequency of the test cue should have no effect. Instead, normatively determined cue-to-target strength should be the only variable to affect recall. Experiment 2 held cue-to-target strength constant and crossed test cue frequency with target frequency for reasons that will be detailed later.

NORMATIVE STUDY

Method

Materials. The stimuli consisted of 5,019 words, including 3,814 nouns, 652 adjectives, 351 verbs, and 202 adverbs and other parts of speech. These items were normed over a 20-year period, collected from more than 6,000 participants, and are currently available on the Web (Nelson, McEvoy, & Schreiber, 1998). Reliability estimates, as well as our reasons for collecting these norms, for selecting particular words to norm, and for using the free association procedure, have been described more fully elsewhere (Nelson, McEvoy, & Schreiber, 1998; Nelson & Schreiber, 1992). In the latter paper, a preliminary report that focused on concreteness and set size and that relied on about half of the current sample of words can be found.

Participants. The participants were undergraduates in introductory psychology at the University of South Florida who took part on a voluntary basis on the 1st day of class. An average of 149 ($SD = 15$) participants responded to each stimulus.

Procedure and measures. All participants were given discrete association instructions to produce the first word coming to mind that was meaningfully related or strongly associated to each stimulus. They were given a booklet containing instructions on the initial page and the stimulus words on the remaining pages. An underlined space was placed next to each word, and participants were asked to write their response in this space. The order of the pages was unsystematically randomized for each person, no one was required to respond to more than 120 words, and we attempted to prevent related words from appearing in the same booklet.

Frequency counts (FREQ) of the number of times a given word occurs in a sample of 1 million printed words were taken from Kučera and Francis (1967). FREQ averaged 78 per million for the words normed in the database ($SD = 389$, range = 0–9999). FREQ values were transformed using the formula $\log(1 + \text{frequency})$ for all analyses to better approximate a normal distribution (a “1” was added to each value because the log of a frequency value of zero is indeterminate). Accessibility counts (ACCESS) were determined by a computer program that simply counted the number of times that each normed word was produced as a response by another normed word. Each word was produced by an average of 12.68 other words ($SD = 20.87$, range = 0–324/5,019). The assumption was that words that are produced more often as responses, such as FOOD, MONEY, and CAR, are more accessible in memory (see also the availability measure reported by Rubin & Friendly, 1986). This index was transformed using the same algorithm as was used for the printed frequency index: $\log(1 + \text{accessibility})$. Another index of accessibility, what we call probability of resonance (PR), calculates the probability that a word’s associates produce it as an associate. For some words, each of their associates also produces the word as an associate (e.g., BEE), whereas for others, none of them do (e.g., BREED). ACCESS provides an index of all words in the normative database that produce a particular word as an associate, whereas PR is limited to those words that are also associates of that particular word. PR represents a subset of ACCESS, and the two measures should be correlated.

The “forward strength” (FSG) of any particular response or associate was inferred from the number of participants producing it in the sample. For example, 114 participants were asked to free associate to DINNER, and 61 and 13 people, respectively, produced *supper* and *eat* as responses. The respective probabilities of these responses are .535 (61/114) and .114 (13/114), and we infer that the DINNER to *supper* connection has a higher level of FSG than the DINNER to *eat* connection.

Meaning set size (MSS) is a normally distributed variable that indexes the number of associates linked to each normed word. MSS is estimated by counting the total number of different responses given by 2 or more participants in each sample (Nelson & Schreiber, 1992;

Nelson, Schreiber, & McEvoy, 1992). Idiosyncratic responses are excluded because they are unreliable, but it should be noted that most of these responses are not “off the wall” but appear to be weakly related associates (Nelson & Schreiber, 1992). The MSS index is highly correlated with the total number of different responses, including idiosyncratic responses ($r = .79$, $n = 5,019$), but using the more inclusive measure did not appreciably alter any conclusions reported in this article.

Mean connectivity (MC) among the associates of each normed word was calculated for 4,097 words in the sample. These values were determined by collecting additional discrete association norms for each of the associates of the normed words, with each associate being normed by a different group of participants (e.g., Nelson, Bennett, Gee, Schreiber, & McKinney, 1993; Nelson, McKinney, Gee, & Janczura, 1998). An $n \times n$ matrix was then constructed and the mean number of connections among the associates within each set was calculated. For example, the word DINNER has an MSS value of 5 associates, including *supper*, *eat*, *lunch*, *food*, and *meal*, and there are an average of 3.40 connections among them (17 connections/5). The distribution for MC tends to be somewhat more peaked and slightly skewed to the right than a normal distribution.

Concreteness ratings (CON) were obtained from available norms whenever possible (Paivio et al., 1968; Toglia & Battig, 1978). When such ratings were unavailable, they were collected using instructions provided by Spreen and Schulz (1966). Participants were told that any word referring to objects, materials, or persons should receive a high concreteness rating and any word referring to an abstract concept that cannot be experienced by the senses should receive a low concreteness rating. The scale ranged from 1 to 7, with a 1 representing “highly abstract” and a 7 representing “highly concrete.” No fewer than 40 individuals participated in the concreteness rating task, and participants never rated more than 150 words. Such ratings were available for 3,260 words, and they show a clear bimodal distribution (Nelson & Schreiber, 1992).

Relationships among these indices. Correlations among FREQ, ACCESS, FSG, MSS, MC, and CON are presented in Table 1. Sample size is also reported for each correlation because we did not have concreteness ratings and mean connectivity values for all of the items. It also varies as a function of how an index is measured. With the exception of FSG, the basis for the remaining indices consisted of the pool of 5,019 normed words. Because pairs of words form the basis for calculating FSG, the pool consisted of the 72,176 pairs produced by each normed word and each of its responses. Limiting the database to nouns appeared to have only minimal effects on these correlations, so we pooled the data across parts of speech. Furthermore, although all of the correlations among the variables are reported, we will describe and discuss only those values that are related to frequency because the present paper is focused on this variable. Hence, the key information is provided in the first column of the table.

Frequency, concreteness, and associative organization. Table 1 indicates that frequency is negatively related to concreteness ($r = -.29$). More common words tend to be more abstract than less common words, and this finding is consistent with previous results, but the magnitude of the relationship is stronger in the present sample (Rubin & Friendly, 1986, $r = -.06$; but see Paivio et al., 1968, who reported a small positive correlation, $r = .12$). Table 1 also shows that frequency has a very slight but negative relationship with associative organization as indexed by the mean number of connections among a word’s associates ($r = -.06$). Although more common words tend to have fewer connections among their associates, the relation is so weak that, practically speaking, it can reasonably be ignored. Although high-frequency words tend to be more abstract, the degree of organization within a word’s associates appears for the most part to be independent of frequency of occurrence.

Frequency as accessibility from other words. As can be seen, the frequency of the word being normed was moderately correlated

Table 1
Correlation Matrix for Log (1 + Frequency) (FREQ),
Log (1 + Accessibility) (ACCESS), Probability of a
Resonant Connection (PR), Concreteness (CON), Mean Associative
Connectivity (MC), Meaning Set Size (MSS), and Forward Strength (FSG)

	FREQ	ACCESS	PR	CON	MC	MSS
ACCESS	.53 (5,019)					
PR	.34 (4,097)	.73 (4,097)				
CON	-.29 (3,260)	.02 (3,260)	.10 (2,927)			
MC	-.06 (4,097)	.03 (4,097)	.04 (4,097)	-.01 (2,927)		
MSS	.15 (5,019)	.12 (5,019)	-.21 (4,097)	-.13 (3,260)	.20 (4,097)	
FSG	-.04 (72,176)	-.02 (72,176)	.07 (57,079)	.05 (45,751)	-.04 (57,079)	-.25 (72,176)

Note—Sample sizes are shown in parentheses below each correlation and they differ because all values were not available for all items and because the basis of the correlations was either individual items or pairs.

with access ($r = .53$). This finding is consistent in both direction and magnitude with previous results based on smaller samples of items (e.g., Howes, 1957; Rubin & Friendly, 1986). PR was highly correlated with ACCESS, as expected ($r = .73$), and was moderately correlated with printed frequency ($r = .34$). High-frequency words would appear to be more accessible as responses from other words, as has been suggested (e.g., Howes, 1957; Rubin & Friendly, 1986; Underwood & Schulz, 1960). The accessibility findings suggest that common words tend to be more likely to be produced as responses by other words in general and by their own associates in particular.

The dimensions thus far considered were only weakly related to the probability of a nonresponse (NR), defined as failing to respond in the free association task. The average number of NRs was .03 ($SD = .04$, range = 0–.78). The correlations between NR and CON (–.22), MC (–.08), FREQ (–.08), accessibility (–.17), and PR (–.17) were generally low and negative. Words having higher values on each of these dimensions were slightly less likely to be associated with response failure. We speculate that such failures are more likely to occur when the word is poorly understood (e.g., NR = .78 for USURP) and when words are experienced with such high frequency that they begin serving more like function words (e.g., NR = .15 for USE).

Frequency as accessibility to other words. Table 1 shows that printed frequency is only very weakly correlated with a word's set size ($r = .15$). A fine-grain plot of set size as a function of word frequency showed that words with frequencies of zero ($n = 339$), 1–5 ($n = 1,211$), and 6–10 ($n = 597$) had MSSs of 12.54 ($SD = 4.91$), 13.32 ($SD = 5.11$), and 14.27 ($SD = 4.96$), respectively. Frequency intervals above these values produced set sizes that fluctuated by only an item or so, with an average MSS value of 14.38 ($SD = 5.15$) for the 5,019-word pool. Hence, words having very low frequencies of 0–5 per million generally showed smaller sets, but the difference approximates only a single item on average. Uncommon words appeared to have about as many associates as more common words. This conclusion differed from that of Postman (1970), who collected discrete as well as continuous association norms and found that, in the discrete task, frequency and set size were related by a u-shaped function. Set size decreased from low to moderate levels of frequency and then increased at the highest frequency level. This discrepancy, however, may have been related to using only 24 items at each level of frequency.

The correlation between the frequency of the normed word and the FSG of each of its associates was both negative and weak ($r = -.04$). In the free association procedure, there appears to be no relationship between printed frequency and response strength. Nevertheless, it seemed plausible to assume that the strongest responses might show an association between frequency and strength that is overshadowed when each of the associates of the normed words is included in the pool. To examine this possibility, the correlation between frequency and the strength of the primary associate was calculated for the word pool. This correlation was negative and only slightly higher ($r = -.07$, $n = 5,019$). Hence, regardless of whether responses of all strengths or only the primary associate is considered, responses produced to high-frequency words tend to be emitted with about the same strengths as those produced to low-frequency words.

Finally, we determined whether there was any relationship between the frequency of the word being normed and the frequency of the responses produced. High-frequency words may be more likely to produce high-frequency responses and low-frequency words may be more likely to produce low-frequency responses. Such a correlation would seem natural because of a history of co-occurrence in various samples of text. In this instance, a weak positive relationship was obtained ($r = .20$, $n = 72,176$).

Discussion

The present normative study appears to provide a clear picture of the correlational features of printed frequency. Common words, generally speaking, tend to be more abstract but their associates seem to be no more organized than the associates of less common words. In addition, low-frequency stimulus words do not seem to produce a greater sampling of high-frequency words (Gillund & Shiffrin, 1984). There is instead a weak tendency for frequency matching in response production, with high-frequency words showing a weak tendency to sample high-frequency words and low-frequency words producing a weak tendency to sample low-frequency words. More importantly, by comparison with less common words, more common words appear to have about the same number of associates and these associates are produced at about the

same levels of strength. Positive evidence showing that more common words tend to be connected to more other words or that they have stronger connections to other words was clearly lacking. In contrast, positive evidence showing that more common words tend to have more connections from other words did emerge. High-frequency words are produced by more words in general, as has been found in other studies (e.g., Rubin & Friendly, 1986), and they are more likely to be produced by their own associates. Hence, the evidence favors the conclusion that more common words are more accessible in memory because more words in the language bring them to mind. This conclusion is consistent with the idea that printed frequency norms tend to capture more about processes involved in writing text than about the influences of reading text. It is also consistent with the conclusion that the effects of printed frequency in cognitive tasks should be attributed to differences in memory accessibility rather than to the existence of connections to more words or to stronger connections to other words.

EXPERIMENT 1

The results of the normative study are interesting and useful primarily because such work tells us about the magnitude of relationships among various indices, knowledge that can be used to help design and interpret the results of experiments concerning frequency and other word attributes as well. For example, they suggest that a researcher interested in the effects of frequency on recognition should be mindful of confounding frequency with concreteness because these variables are correlated. They also suggest that, contrary to what has been assumed in the SAM model (Gillund & Shiffrin, 1984), high-frequency targets and distractors are unlikely to have stronger connections to a set of recently studied words in comparison with low-frequency words. SAM assumes that high-frequency words have stronger connections, what we have called the frequency–strength hypothesis, but this hypothesis leads to an expectation of a positive frequency–strength correlation that did not emerge in the free associations involving more than 72,000 pairs of words.

The absence of such a correlation, however, is not particularly troublesome for SAM because the model assumes that the absolute strength is what is important. Gillund and Shiffrin (1984) reasonably argued that free association norms are likely to be misleading because such norms index relative strength, not absolute strength (p. 44). As they suggested, the HORSE–STIRRUP connection may be stronger in an absolute sense than the STIRRUP–HORSE connection even though free association norms would be likely to show that the opposite was true. Free association norms may lead us to a false conclusion because HORSE, as a high-frequency word, is likely to be connected to many more words than STIRRUP, which is a low-frequency word. The HORSE–STIRRUP connection seems weaker only because HORSE activates so many related words in addition to STIRRUP.

This interpretation of norms assumes that more common words are connected to more words, an assumption that the present findings do not support. The absolute strength interpretation of the frequency–strength hypothesis is persuasive but the normative findings nag allegiance, particularly findings indicating that, contrary to expectation, common words seem to produce about the same number of close associates as less common words. If, for the sake of argument, HORSE and STIRRUP have the same number of associates, it cannot be said that number of associates deceives us about the connection strengths between these items. The normative findings, however, are correlative in nature and therefore are suspicious in the eyes of many experimentalists, ourselves included. When variables involve manipulations of the materials to be learned or remembered, careful experiments are needed to demonstrate how the attributes behave when isolated from correlated attributes. For example, previous research has shown that the number of associates linked to a studied word has a negative effect on recall in an extralist cuing task (e.g., Nelson et al., 1992). With other factors controlled, studied words having more associates as indexed by free association norms were less likely to be recalled. More importantly for present purposes, the influence of number of associates was shown to have additive effects with concreteness (Nelson & Schreiber, 1992), connectivity among the associates (Nelson et al., 1993), and printed word frequency (Nelson & Xu, 1995). In each of these demonstrations, the norms indicated that number of associates and each of the other attributes were weakly correlated at most and therefore likely to have functionally independent effects on recall. The free association norms predicted several relationships that were confirmed through experimentation.

The purpose of Experiment 1 was to test the frequency–strength hypothesis in a simple experiment by selecting different types of words to be used as test cues for recently studied words. As suggested earlier, extralist cued recall can be used to provide another index of strength of connection besides free association. In Experiment 1, with other attributes held constant, the recall of studied words was prompted by four different types of cues created by crossing cue frequency with cue-to-target strength. Test cues were either common or uncommon words and, as measured by free association norms, they were either strongly or weakly related to the studied words. According to the strongest interpretation of the frequency–strength hypothesis, high-frequency cues will be more effective than low-frequency cues and, because it represents a misleading index of strength, normatively measured cue-to-target strength will have no effect. For example, although equally strong cues according to the norms, because BOX occurs more frequently than CUBE, it should be a more effective test cue for the studied word SQUARE. In contrast, according to what might be called the norms–strength hypothesis, recall will be affected by cue-to-target strength but not by cue frequency, which is presumably uncorrelated with strength. Given that pre-

vious studies have shown that strength has a substantial influence on probability of cued recall (e.g., Bahrick, 1970; Nelson & McEvoy, 1979; Nelson et al., 1992), both variables may influence performance. In this case, a weaker interpretation of the frequency–strength hypothesis prevails. Free association norms may correctly index some aspects of the strength of an association, but for some unknown reason it may be insensitive to the frequency–strength component.

Method

Design and participants. The experimental design formed a 2×2 within-subjects factorial with test cue frequency (high, low) and cue-to-target strength (strong, weak) as the principal independent variables. Twenty-four participants served in the experiment for course credit with 6 people randomly assigned to each of four different lists.

Materials. The normative database was used to select 24 target words to serve as the study list and 4 words to serve as test cues for each of these studied items. This procedure resulted in four different lists of cue–target pairs; all materials are presented in Appendix A. High-frequency test cues occurred an average of 160 times per million words ($SD = 146$) and low-frequency words occurred an average of 7 times per million ($SD = 6$). Test cues with strong forward connections to their targets produced them with a probability of .26 ($SD = .06$) in free association and weakly related cues produced them with a probability of .06 ($SD = .03$).

Cue frequency and strength were carefully controlled at each level of the other variable. Other variables known to affect probability of cued recall were also controlled (see Nelson, Bennett, & Leibert, 1997, and Nelson, McKinney, et al., 1998, for details). Backward strength from target to cue averaged .04 ($SD = .10$), shared associate strength averaged .02 ($SD = .03$), and mediated strength averaged .04 ($SD = .06$). Both shared associate and mediated strength refer to two-step indirect connections between the test cue and target. For shared associates, both test cue and target share one or more common associates (e.g., both *poet* and *author* produce *book* as an associate). Shared associate strength is calculated by cross-multiplying normatively measured cue-to-associate strength by target-to-associate strength for each common associate and summing the results. Mediators involve an associate that provides an indirect link between the cue and target (e.g., *poet* to *writer* to *author*). Mediated strength is calculated by cross-multiplying the strengths of cue-to-mediator and mediator-to-target links and summing the results for multiple mediators. Finally, each test cue produced an average of 13.56 associates in free association ($SD = 4.72$), so the meaning set size of the test cues was close to the mean of the set size distribution.

Because the manipulations of test cue frequency and forward strength were created by selecting different types of cues, the same words served as target words for all cue types. The advantage of this design is that the same set of words was being recalled across all cue types and any differences in results cannot be attributed to target differences. Nevertheless, target characteristics known to affect cued recall were held constant within each frequency–strength condition. When pooled over conditions, meaning set size of the target, frequency, concreteness, connectivity, and probability of a resonant connection averaged, respectively, 15.70 ($SD = 6.15$), 194 ($SD = 200$), 3.55 ($SD = 1.28$), 1.73 ($SD = .68$), and .72 ($SD = .12$). The target words generally consisted of frequently occurring, middle-of-the-concreteness scale words having about an average number of associates. Although their associates were only moderately connected with each other, most of them produced the target as an associate.

Procedure. Participants served in individual sessions and received a single study–test trial. During study, the target words were presented in uppercase letters on a Macintosh computer at a 3-sec rate, and all words were read aloud by the participant when shown. Participants were asked to remember as many targets as possible without being told about how they would be tested. Particular learning strategies were not suggested. Just before showing the study list, all participants were shown a short list of people's names to familiarize them with the presentation rate.

The test instructions were read to the participants immediately following the last study presentation. These instructions indicated that a meaningfully related test cue would be shown to help them recall each word in the list, that the cue should first be read aloud, followed by an attempt to recall the list word. Clarifying examples were included. This test trial was self-paced, and guessing was encouraged but not required. Finally, the order of the test cues and the study words was unsystematically randomized for each participant.

Results and Discussion

The probabilities of correct cued recall when test cue frequency was high were .59 and .40, respectively, for strong and weak cues; when test cue frequency was low, the respective probabilities were .59 and .36 for strong and weak cues. Hence, recall appeared to vary with cue-to-target strength but not with the frequency of the test cue. Recall was higher when strong cues were used (.59) than when weak ones were used (.38). In contrast, high-frequency cues (.50) were only slightly more effective than low-frequency cues (.48). An analysis of variance (ANOVA) of these data indicated that the strength effect was reliable [$F(1,23) = 24.10$, $MS_e = .045$] and that neither test cue frequency ($F < 1$, $MS_e = .038$) nor the interaction ($F < 1$, $MS_e = .033$) was significant.

These findings are inconsistent with expectations derived from both strong and weak versions of the frequency–strength hypothesis. Test cue frequency failed to have reliable effects for either normatively defined strong cues or for normatively defined weak cues. Recall was higher for stronger cues, and low-frequency cues were just as effective as high-frequency cues in prompting the recall of the same high-frequency targets. This finding suggests that more common words generally are not more strongly connected to their associates than less common words, as has been suggested. The finding also implies that free association norms provide valid estimates of the strengths of connections between related words. Normatively strong and weak cues performed as expected.

EXPERIMENT 2

Experiment 2 was undertaken as an extended replication of the first experiment using different items. Test cue frequency was varied but instead of crossing this variable with forward cue-to-target strength, it was crossed with target frequency. With forward strength held constant, cue frequency was either high or low and target frequency was either high or low. The purpose of this experiment was to evaluate a variation of the frequency–strength hypothesis. Although it seemed unlikely, this hy-

pothesis may be correct when applied to low-frequency targets. High-frequency cues may be more strongly related to their associates than low-frequency cues only when these associates are low in frequency. In the extralist cuing task, test cue frequency may affect recall only when the retrieval of low-frequency words are involved. Put another way, free association may provide an invalid indicator of forward strength when applied to infrequently appearing responses. If so, Experiment 2 will show that high test cue frequency will benefit cued recall only when target frequency is relatively low. Alternatively, if free association provides a valid indication of forward strength, cue frequency should once again fail to influence recall.

Method

Design and participants. The experimental design formed a 2×2 within-subjects factorial varying test cue frequency (high, low) and target frequency (high, low). Thirty participants served in the experiment for course credit with 15 randomly assigned to one list and 15 to the other.

Materials. The norms provided the necessary data for constructing two 24-item lists, which are shown in Appendix B. Test cue frequency averaged 115 ($SD = 42$) and 5 ($SD = 4$), respectively, for high- and low-frequency cues. Target frequency averaged 140 ($SD = 42$) and 5 ($SD = 3$). Target frequency was equated at each level of cue frequency.

Other variables were controlled within each condition of cue-target frequency. Forward, backward, overlap, and mediated strengths averaged, respectively, .22 ($SD = .06$), .11 ($SD = .11$), .03 ($SD = .04$), and .02 ($SD = .03$). Generally speaking, the cues and targets were strongly related. The cues and targets had average-sized sets in terms of the normative sample with means of 14.35 ($SD = 3.47$) and 14.56 ($SD = 4.64$), respectively. In addition, target concreteness, connectivity, and probability of a resonant connection were also controlled, averaging 4.77 ($SD = 1.55$), 1.65 ($SD = .79$), and .46 ($SD = .19$).

Procedure. The procedures for extralist cued recall were the same as those used for Experiment 1, including the use of a single study-test trial of individual items, a self-paced test, and so on.

Results and Discussion

The probabilities of cued recall when high-frequency targets were being recovered were .61 and .59, respectively, for high- and low-frequency cues. When low-frequency targets were being recovered, the probabilities for high- and low-frequency cues were .69 and .72. From these values it was clear that test cue frequency had no apparent effect. Low-frequency cues (.65) produced the same level of recall as high-frequency cues (.66). In contrast, low-frequency targets (.71) were more likely to be recalled than high-frequency targets (.60). The results of a repeated measures ANOVA indicated that the effect of target frequency was significant [$F(1,29) = 15.62$, $MS_e = .026$]. Neither test cue frequency ($F < 1$, $MS_e = .017$) nor the interaction ($F < 1$, $MS_e = .034$) was reliable.

Once again, cue frequency effects were not found. Uncommon words were just as effective as common words as test cues and, on this occasion, this result was observed for targets that were high in frequency and for those that were low in frequency. These findings are inconsistent with the hypothesis that high-frequency words are more strongly connected to other words. However, the conclusion that the cue frequency-strength hypothesis must be

rejected would be based on weak grounds for two reasons. First, the basis of this conclusion is a null result, and second, the findings are derived from a relatively small sample of several hundred associatively related pairs of words among the many thousands that are possible. Null results are difficult to interpret and the findings may not generalize to other pairings. We attempted to alleviate both concerns by adding the results of the present experiments to the results of 24 other extralist cued recall studies and subjecting the result to a multiple regression analysis. These studies included 2,152 pairs of associatively related words selected and studied for a variety of reasons, including work on concreteness, target frequency, study instructions, and so on (the findings will be described in full in a separate report). An average of 10.62 ($SD = 3.19$, range = 8–24) participants studied each target of the pair and then received its cue during test. Probability of correct recall served as the dependent variable, and both cue and target frequency (using the log of $1 + \text{frequency}$) served as the independent variables. The results showed that these variables were significantly associated with recall [$r = .16$, $F(2,2149) = 27.78$, $MS_{\text{res}} = .067$]. For cue and target frequency, the t values and standardized β coefficients were, respectively, $t = -.66$ ($-.015$) and $t = -.687$ ($-.154$). Both variables were negatively related to recall, but test cue frequency was not significantly related. Only target frequency reliably predicted recall. Uncommon words were more likely to be recalled than common words. This finding replicates the results of several other studies and has been attributed to a fast, efficient recognition process that follows a sampling process in extralist cued recall (Gillund & Shiffrin, 1984; Nelson & Xu, 1995). People are asked to use the test cues to recall words just studied so that recovered words, or at least some portion of them, are checked for list membership before being reported.

GENERAL DISCUSSION

What do printed frequency norms measure? This question provided the impetus for pursuing the present research and the answer was sought in correlational explorations of free association norms and in experimental manipulations within the extralist cued recall task. Within the limits of these tasks, the findings appear to be straightforward. The free association task was based on data from 5,019 words with associates obtained from more than 6,000 participants, and it indicated that more common words show some tendency to be more abstract than do less common words. However, such words appear to have about the same number of associates as less common words and the average number of connections among the associates of common and uncommon words appears to be about the same. High-frequency words also produce about the same numbers of nonresponses.

High-frequency words also do not appear to be more strongly connected to their associates than low-frequency words. The correlation between test cue frequency and forward strength was very small and in the opposite di-

rection. More importantly, the results of the cued recall task indicate that, with all known variables being equal, high-frequency test cues held no advantage over low-frequency test cues. As extralist retrieval cues, low-frequency words were just as effective as high-frequency words. Hence, even though, by definition, high-frequency words are experienced more often in print, other than tending to be somewhat more abstract, they seem to differ little in item-information content and in the strength of connection *to* other related words. Nevertheless, higher frequency words differed in an important way. Their associates are more likely to produce them as associates and, in general, more words produce them as associates. High-frequency words appear to be more accessible as responses from a greater variety of other words, a result that others have also observed (Howes, 1957; Johnson, 1956; Rubin & Friendly, 1986; Underwood & Schulz, 1960, p. 88).

Implications

The present findings have implications for what printed frequency counts measure, for what free association norms measure, and for theories of recognition. Each of these implications will be considered in turn.

Implications for interpreting frequency counts.

Frequency counts are correlated with concreteness and accessibility. The correlation with concreteness suggests that researchers interested in printed frequency as an index of word retrievability must be mindful of its association with concreteness and take care to control it or to manipulate and explore the results of both variables. The correlation with accessibility would appear to validate what is generally believed about such norms: Words that appear more often in print are better known and more readily retrieved for use in various tasks. On average, they are experienced in more contexts, more recently, and more words produce them as associates. More common words have more connections from other words and therefore tend to be selected ahead of other words that might be used to convey the same message in discourse.

More common words, however, do not seem to have more or stronger connections *to* other words. Although the early literature suggested that common words have slightly more associates (e.g., Noble, 1952), problems of small item samples and the use of the continuous rather than the discrete association task leave room to doubt the early conclusions. In continuous association, participants produce as many words as possible in a fixed period of time to a single cue, and it is difficult to know how response production might be influenced by response chaining and retrieval inhibition as frequency and other item features are varied.

Response chaining and retrieval inhibition were circumvented in the present normative study because free discrete association required participants to produce only the first related word to come to mind. Because these potential problems were avoided and because a large word sample was employed, we believe that it is now more reasonable to conclude that, though more common words have more connections *from* other words, they have about

the same number of connections *to* other words. This conclusion could mean that printed frequency counts may be more representative of how writers write than what readers read. If such counts reflected only what readers read, it would seem natural to presume that common words would have more connections to and more connections from other words because they appear in a greater variety of contexts and therefore with a greater variety of words. In contrast, the writer way of thinking about what frequency counts measure suggests that these counts index the accessibility of the words that writers use in producing meaningful discourse. Of course, what may be incorrect here is the assumption that reading strengthens both from and to connections. In their attempt to comprehend the text, readers may rely on the same accessibility factors that control the writer's choice of words. Readers may read with the writer's intent in mind and, as they plow through the text, previously occurring words and ideas constrain words that follow. If readers and writers act in consort in the common cause of understanding, the distinction between writer and reader points of view vanishes. What may also need to vanish is the tendency to believe and act as if more common words are connected to more words and have stronger connections to these words.

Implications for free association norms. In discrete free association, people typically use a word to recover the first related word they think of, and the strength of any given response is inferred from the number of people in the sample that produce it. For example, if the probability of one response is .30 and the probability of another is .10, we can say that the first response is stronger or more probable than the second. As long as the responses have been produced by two or more people in the same sample, we can also say that this difference is reliable (e.g., Nelson & Schreiber, 1992). We can also say that these indices are generally valid in the sense that they predict performance in many different tasks (Canas, 1990; see also Cramer, 1968, for a review). However, we cannot say with any assurance that a .30 response is three times stronger than a .10 response because the scale underlying the measurement is unknown. Given the reliability and apparent validity of the measurement, it would seem reasonable to conclude that, at the minimum, an ordinal scale has been achieved. Ordinal metrics permit "greater than" and "less than" conclusions, but they do not allow conclusions about the magnitude of a difference. Conclusions about absolute magnitude are allowed only when a ratio scale has been achieved, and we can say that the strength of one response is, say, three times greater than another.

Given this apparent limitation in the scale, an argument that free association norms are misleading because the strength metric could mean something different for words having different characteristics becomes immediately credible. For instance, the frequency-strength hypothesis provided the primary motivation for this article because it seemed compelling and we thought that someone should attempt to confirm it (Gillund & Shiffrin, 1984). However, because the scale underlying the strength

metric is unknown, the finding of a very low correlation between frequency and forward strength in free association provides an insufficient challenge to this hypothesis. As suggested, the HORSE to STIRRUP connection may be stronger in an absolute sense than the reverse connections even though STIRRUP is much more likely to produce HORSE in free association.

What was needed was another measure of the strength of a connection that was potentially independent of free association. In the present case, such a measure was provided by using words as extralist retrieval cues for prompting the retrieval of recently studied words. Extralist cues are unavailable during study so their effectiveness as cues is unlikely to depend on connections formed during this period. At the outset of the experimental enterprise, it was unclear whether the effectiveness of such cues would be better predicted by test cue frequency or by free association strength. That strength prevailed over frequency suggests that the norms were not only not misleading but also that they were predictive of memory recovery (see also Bahrick, 1970). This result substantiates the validity of the norms as a means for indexing preexisting strength for both high- and low-frequency words. In a regression study currently under way that is based on many experiments from our lab, free association measurements explain 28% of variance in the extralist cued recall task, with forward strength alone explaining half of this variance.

With this accomplishment we are tempted to conclude that free association norms directly assess strength about as well as it can be measured. Succumbing to such a temptation, however, would be premature in the absence of knowledge about the measurement scale but also because we would be drawing inferences in the absence of a theory of free association. The recent history of free association norms has been to use them only to measure preexisting lexical knowledge. For example, in work on cued recall with various types of cues researchers have relied heavily on using free association norms for meaning cues, taxonomic categories, rhymes, as well as stem and fragment cues. For the most part this reliance grew out of serious attempts to control preexisting forward strength so that the differential effectiveness of various types of cues could be compared. This use of such norms follows a long and honorable measure-then-manipulate tradition in psychological research, but it treats free association as a yardstick instead of an important memory task that is worthy of being understood in and of itself.

There is no reason why free association cannot serve in both roles, for one is likely to inform the other. For starters, we suggest that free association shares two processes with cued recall (Nelson, McKinney, et al., 1998). First, we assume that, as in cued recall, reading a familiar word produces a parallel activation of its associates in memory. This activation guarantees a pool of related items. Second, we assume that items in this pool are sampled according to their preexisting strengths under the constraints of the task. In extralist cued recall, the constraints are substantial because only a single instance in the pool is generally defined as acceptable (i.e., the item

must have appeared in a study list). Because of this constraint, the forward connection between the initiating cue and the studied item is important. In addition, the item-specific characteristics of this item and its other associative relationships with the cue and with the other items in the pool also become important. In contrast, in free association any item in the pool is acceptable. No known experimental constraints are present. The strength of the forward connection between the initiating cue and each of the items in the pool is of singular importance. Stronger items are more likely to be sampled than weaker items, with strength determined by prior experience.

This framework for free association is sketchy because it sidesteps the representational issue for the moment, but it serves several functions. First, it emphasizes the need to develop more complete models of this process. Second, as simple as it is, it helps to explain why measures taken from free association norms successfully predict extralist cued recall as well as other performance. Third, along with the findings in this paper, this framework suggests that free association may provide the best index of forward cue-to-target strength developed to date for identifying the strongest associates of a given word. No other procedure is as likely to be free of external constraints. For example, co-occurrence norms are obtained from large samples of text, but the very nature of text imposes semantic constraints that influence co-occurrence frequency. Finally, the framework provides some groundwork for experimentation with free association. As an illustration, it suggests that placing constraints on free association such as requiring participants to produce an associate with a specific letter will alter the index of cue-to-target strength. Using free association as a measurement tool has produced a systematic body of evidence that must be explained by memory models and further experimentation with the free association task itself is likely to produce other important findings as well.

Implications for recognition models. Our free association findings suggest that high- and low-frequency words have approximately the same numbers of associates, that these associates are of approximately equal strength, and that there is a weak tendency for frequency matching in sampling responses. The latter finding implies that, in the aggregate, low-frequency cues such as STIRRUP are more likely to produce low- than high-frequency responses. Taken together, such findings suggest that high-frequency words are not any stronger than low-frequency words as cues for recovering other words (see Gillund & Shiffrin, 1984, p. 34). For example, the low-frequency word STIRRUP does not seem to cause the sampling of HORSE, a higher frequency word, because other competing associations are so few or so weak. HORSE is sampled because these words are strongly associated as a result of experience. More important, as the experiments on test cue frequency show, high-frequency cues are no more effective than low-frequency cues in recovering list words. Put simply, low- and high-frequency words are about equally related to other words. This aspect of the frequency-strength hypothesis appears to be

false, and this conclusion is inconsistent with an important assumption of SAM. In particular, this conclusion is contrary to the assumption that high-frequency distractors are more strongly connected to list words than are low-frequency distractors (the assumption that $d_H > d_L$).

After the research on the present project was completed, Shiffrin and Steyvers (1997) revised SAM because of difficulties encountered in explaining list strength and mirror effects. The revision (REM) is a similarity model in which words are represented as vectors of feature values that are subjected to a matching process during episodic recognition tests. The frequency effect (p. 153) is explained by assuming that the representations for high-frequency words are generated to have more common values and therefore less differentiating features than low-frequency words ($g_h > g_l$). As a result, when a target or distractor is matched to the list, the matching feature values for high-frequency words tend to be lower and contribute less discriminating information about the target. In signal detection terms, d' is less for high- than for low-frequency words because their features are less diagnostic. If strength of connection from distractors to list words can serve as a differentiating feature in this model, then REM, like SAM, would also attribute the frequency effect, at least to some extent, to the hypothesis that high-frequency distractors are more strongly connected to list words than are low-frequency distractors.

The present findings are inconsistent with the frequency–strength hypothesis and therefore with SAM and possibly REM, but a caveat is in order because the findings are based on directly associated words. In contrast, in the typical recognition experiment, directly associated words are purged from the lists. Because experimenter judgments rather than norms are used to screen items, the success of this procedure is unknown, but the pairs are probably more weakly related on average than the ones used in the present studies. The frequency–strength hypothesis may hold for pairs having connections that are too weak to appear as direct associates in free association. For example, it may hold for pairs that are associated indirectly through other words. We examined this possibility in a sampling study by randomly selecting 200 high- and 200 low-frequency nouns from the normative database and forming two frequency-based lists. High- and low-frequency words occurred 237 (range = 119–1,207) and .25 (range = 0–1) times per million words, respectively. Homographs, proper names, and dual words (e.g., *peanut butter*) were eliminated, and aside from these restrictions, no attempt was made to screen related words in either list at this point. Instead, each list was run separately through the ListChecker program (Schreiber, 1993), which made all pairwise comparisons among the words of each list and used the normative database to report indirect as well as direct connections.

The results of the sampling study showed two interesting trends regarding the number and the strength of connections. First, when words having direct connections were purged (e.g., *husband–wife*), each high-frequency

word was indirectly connected to an average of 45.94 other high-frequency words in the list (out of 177 words). In contrast, each low-frequency word was indirectly connected to an average of only 19.44 low-frequency words (out of 191 words). That is, 26% of the high-frequency pairs had at least one indirect connection involving 2-step mediators or shared associates, whereas only 10% of the low-frequency words had such connections. However, given the presence of an indirect connection, the number of such connections within each pair did not differ appreciably as a function of either frequency or type, averaging 1.24 connections per pair. Second, and what is more important for the frequency–strength hypothesis, indirect strength did not vary with frequency. Mediated strength per pair averaged .003 for both high- and low-frequency words, and shared associate strength averaged .004 for high- and .011 for low-frequency words. The values for number and strength of the indirect connections did not change as a function of frequency when direct connections were unpurged or when they were restricted to low values ($\leq .04$). Finally, the strengths of the direct connections did not vary with frequency, either. Strengths averaged .004 and .001 for high and low frequency for unpurged computations and, given a direct connection, they averaged .02 and .02 when such connections were restricted to $\leq .04$.

The main results of the sampling study showed that, regardless of whether direct connections were screened, there were more indirect connections among the words of the high- than of the low-frequency list, but despite this difference, the average strength of these connections did not vary with frequency. These results of course could be restricted to the specific samples of words used to represent frequency. To evaluate the generality of the relationship between indirect connections and frequency, the normative database of directly related words was used once again, and in this more general case, the correlations were low. The analysis was restricted to pairs having normed responses ($n = 42,202$), and the correlations between printed frequency and number of mediators and number of shared associates were identical at $r = -.01$. For the strength indices, these correlations were also identical at $r = -.06$. These values were unaffected by restricting the computations to pairs with forward strengths $\leq .04$ ($n = 28,397$). Although high-frequency words *can* share more indirect connections in any single sample, it seems unlikely that such sharing is generally true, and, more importantly, neither the single sample used here nor the general case reveals any evidence in support of the frequency–strength hypothesis.

The sampling experiment did show, however, that random selection procedures can produce a list in which there are greater numbers of weak indirect connections among the high- than among the low-frequency words. Hence, some recognition experiments may have confounded the frequency manipulation with the number of indirect connections, and to the extent that such connections can influence recognition, SAM's explanation for

frequency effects would be viable. Nevertheless, given the absence of evidence favoring the frequency–strength hypothesis, it may be time to reconsider other alternatives. For example, the frequency effect in episodic recognition may be more readily explained by focusing on context rather than on word-to-word connections. Early models assumed that recognition was determined by recovering the association of the test item to the context of the encoding (e.g., Anderson & Bower, 1974; Hollingworth, 1913; Tulving, 1976). This task can be conceptualized as a discrimination problem directed toward determining whether the test word seems familiar because it was present in the current context or present in some past context (e.g., see Dennis & Humphreys, 2000). By definition, high-frequency words appear in a greater number of different contexts than low-frequency words and should be disadvantaged when it comes to solving the discrimination problem. People should find it more difficult to determine whether a high-frequency word appeared in the current context or during some past episode because of its appearance in more contexts in a lifetime of learning. One advantage of context models is that they provide a ready explanation for previously established and now replicated findings showing that high-frequency words are produced as associates by more different words. Put simply, the idea of more contexts translates readily into more words. Because of how frequency norms are calculated, number of different contexts is naturally confounded with number of different words. With the proviso that frequency norms index how people choose words when they write text, the implication is that high-frequency words will have more direct connections from other words but not necessarily to other words.

REFERENCES

- ANDERSON, J. R., & BOWER, G. H. (1974). A propositional theory of recognition memory. *Memory & Cognition*, *2*, 406-412.
- BAHRICK, H. P. (1970). Two-phase model for prompted recall. *Psychological Review*, *77*, 215-222.
- CANAS, J. J. (1990). Associative strength effects in the lexical decision task. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *42*, 121-145.
- COFER, C. N., & SHEVITZ, R. (1952). Word association as a function of word frequency. *American Journal of Psychology*, *65*, 75-79.
- CRAMER, P. (1968). *Word association*. New York: Academic Press.
- DENNIS, S. (1995). The Sydney Morning Herald Word Database. *Noetica: Open Forum*, *1*(4).
- DENNIS, S., & HUMPHREYS, M. S. (2000). *The role of context in episodic recognition: The bind cue decide model of episodic memory*. Manuscript submitted for publication.
- GILLUND, G., & SHIFFRIN, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, *91*, 1-67.
- GREGG, V. (1976). Word frequency, recognition and recall. In J. Brown (Ed.), *Recall and recognition* (pp. 183-216). London: Wiley.
- HAYES, J. R., & FLOWER, L. S. (1986). Writing research and the writer. *American Psychologist*, *41*, 1106-1113.
- HOLLINGWORTH, H. C. (1913). Characteristic differences between recall and recognition. *American Journal of Psychology*, *24*, 533-544.
- HOWES, D. A. (1957). On the relation between the probability of a word as an association and in general linguistic usage. *Journal of Abnormal & Social Psychology*, *54*, 75-85.
- JOHNSON, D. M. (1956). Word association and word frequency. *American Journal of Psychology*, *69*, 125-126.
- KÚČERA, H., & FRANCIS, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- NELSON, D. L., BENNETT, D. J., GEE, N. R., SCHREIBER, T. A., & MCKINNEY, V. M. (1993). Implicit memory: Effects of network size and interconnectivity on cued recall. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *19*, 747-764.
- NELSON, D. L., BENNETT, D. J., & LEIBERT, T. W. (1997). One step is not enough: Making better use of association norms to predict cued recall. *Memory & Cognition*, *25*, 785-796.
- NELSON, D. L., & McEVOY, C. L. (1979). Encoding context and set size. *Journal of Experimental Psychology: Human Learning & Memory*, *5*, 292-314.
- NELSON, D. L., McEVOY, C. L., & SCHREIBER, T. A. (1998). *The University of South Florida word association, rhyme, and word fragment norms*. Available <http://www.usf.edu/FreeAssociation/>
- NELSON, D. L., MCKINNEY, V. M., GEE, N. R., & JANCZURA, G. A. (1998). Interpreting the influence of implicitly activated memories on recall and recognition. *Psychological Review*, *105*, 299-324.
- NELSON, D. L., & SCHREIBER, T. A. (1992). Word concreteness and word structure as independent determinants of recall. *Journal of Memory & Language*, *31*, 237-260.
- NELSON, D. L., SCHREIBER, T. A., & McEVOY, C. L. (1992). Processing implicit and explicit representations. *Psychological Review*, *99*, 322-348.
- NELSON, D. L., & XU, J. (1995). Effects of implicit memory on explicit recall: Set size and word frequency effects. *Psychological Research*, *57*, 203-214.
- NOBLE, C. E. (1952). An analysis of meaning. *Psychological Review*, *59*, 421-430.
- PAIVIO, A., YUILLE, J. C., & MADIGAN, S. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monograph Supplement*, *76* (1, Pt. 2), 1-25.
- POSTMAN, L. (1970). The California norms: Association as a function of word frequency. In L. Postman & G. Keppel (Eds.), *Norms of word association* (pp. 241-320). New York: Academic Press.
- ROEDIGER, H. L., III (1974). Inhibiting effects of recall. *Memory & Cognition*, *2*, 261-269.
- RUBIN, D. C., & FRIENDLY, M. (1986). Predicting which words get recalled: Measures of free recall, availability, and pronunciability for 925 nouns. *Memory & Cognition*, *14*, 79-94.
- SALTZ, E. (1967). Thorndike-Lorge frequency and m of stimuli as separate factors in paired-associate learning. *Journal of Experimental Psychology*, *73*, 473-478.
- SCHREIBER, T. A. (1993). *ListChecker* [Unpublished computer program].
- SHIFFRIN, R. M., & STEYVERS, M. (1997). A model for recognition memory: REM—retrieving effectively from memory. *Psychonomic Bulletin & Review*, *4*, 145-166.
- SREEN, O., & SCHULZ, R. W. (1966). Parameters of abstraction, meaningfulness, and pronounceability for 329 nouns. *Journal of Verbal Learning & Verbal Behavior*, *5*, 459-468.
- THORNDIKE, E. L., & LORGE, I. (1944). *The teacher's word book of 30,000 words*. New York: Columbia University, Teachers College Press.
- TOGLIA, M. P., & BATTIG, W. F. (1978). *Handbook of semantic word norms*. Hillsdale, NJ: Erlbaum.
- TULVING, E. (1976). Ecphoric processes in recall and recognition. In J. Brown (Ed.), *Recall and recognition* (pp. 37-73). London: Wiley.
- UNDERWOOD, B. J., & SCHULZ, R. W. (1960). *Meaningfulness and verbal learning*. Chicago: J. B. Lippincott.
- VAN DIJK, T. A., & KINTSCH, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.
- WINNICK, W. A., & KRESSEL, K. (1965). Tachistoscopic recognition thresholds, paired-associate learning and immediate recall as a function of abstractness-concreteness and frequency. *Journal of Experimental Psychology*, *70*, 163-168.

APPENDIX A
Materials for Experiment 1

List 1	List 2	List 3	List 4	Target
Strong-High Frequency Cue	Strong-Low Frequency Cue	Weak-High Frequency Cue	Weak-Low Frequency Cue	
FARM	CAGE	HORSE	ALLIGATOR	ANIMAL
NIGHT	CAVE	EVENING	DUSK	DARK
BEGIN	ORIGINATE	STOP	QUIT	START
AGE	WRINKLE	YOUTH	GRANDMA	OLD
BOX	CUBE	BLOCK	PYRAMID	SQUARE
FIG	LIMB	OUTSIDE	TRUNK	TREE
Strong-Low Frequency Cue	Strong-High Frequency Cue	Weak-Low Frequency Cue	Weak-High Frequency Cue	
JEANS	DRESS	DENIM	CLOTHES	PANTS
DENTIST	PATIENT	SURGERY	OPERATION	DOCTOR
MUG	MEASURE	DISH	COFFEE	CUP
CROOKED	LINE	ARROW	ANGLE	STRAIGHT
INSPECT	SEE	PEER	SAW	LOOK
AD	COLUMN	HOROSCOPE	DAILY	NEWSPAPER
Weak-High Frequency Cue	Weak-Low Frequency Cue	Strong-High Frequency Cue	Strong-Low Frequency Cue	
DANCE	SINGER	RADIO	STEREO	MUSIC
STRANGE	ABNORMAL	SIMILAR	DISSIMILAR	DIFFERENT
NOSE	GARBAGE	SENSE	COLOGNE	SMELL
ARM	LAWN	DESK	STOOL	CHAIR
ANALYSIS	CALCULATE	THOUGHT	RATIONALIZE	THINK
TRUE	FRAUD	TRUTH	DECEIT	LIE
Weak-Low Frequency Cue	Weak-High Frequency Cue	Strong-Low Frequency Cue	Strong-High Frequency Cue	
MUTE	LIBRARY	SHY	SILENCE	QUIET
LEND	SUPPLY	CHARITY	TAKE	GIVE
CLIMB	LEVEL	ELEVATOR	STEP	STAIR
CONQUEROR	LEADER	MONARCH	KINGDOM	KING
SUPERVISOR	INDUSTRY	TOIL	TASK	WORK
BOLT	CLOSED	MAT	KEY	DOOR

(Continued on next page)

APPENDIX B
Materials for Experiment 2

List 1		List 2	
HF Cue	HF Target	HF Cue	HF Target
FAITH	RELIGION	LINE	STRAIGHT
RADIO	MUSIC	TEETH	MOUTH
PATIENT	HOSPITAL	DEGREE	TEMPERATURE
DECISION	CHOICE	VILLAGE	TOWN
GROUND	FLOOR	CONCERN	CARE
BRIDGE	RIVER	IMAGE	PICTURE
LF Cue	HF Target	LF Cue	HF Target
CUBE	SQUARE	ORIGINATE	START
CARTOON	CHARACTER	LAUNDRY	CLOTHES
SIREN	POLICE	CUSTOM	TRADITION
TROPICAL	ISLAND	SPOUSE	HUSBAND
TOOTHBRUSH	TEETH	METEOR	SPACE
LEASE	APARTMENT	BRUNETTE	BROWN
HF Cue	LF Target	HF Cue	LF Target
WINDOW	PANE	ENEMY	FOE
PIECES	PUZZLE	NUMBERS	MATH
SCIENCE	BIOLOGY	STEP	STAIR
FOOT	TOE	OPERATION	SURGERY
PRODUCE	VEGETABLE	CHARACTERISTIC	TRAIT
HOT DOGS	BUN	TERM	SEMESTER
LF Cue	LF Target	LF Cue	LF Target
ADJECTIVE	VERB	SHIELD	ARMOR
FIGHTER	BOXER	NECKLACE	BRACELET
DAM	BEAVER	MOP	BROOM
CHISEL	HAMMER	RAISIN	GRAPE
SCARY	FRIGHT	SELFISH	GREED
CLAM	OYSTER	MOTH	BUTTERFLY

Note—HF, high frequency; LF, low frequency.

(Manuscript received October 15, 1998;
revision accepted for publication June 7, 1999.)