Determinants of the word-frequency effect in recognition memory

BRUCE EARHARD

Dalhousie University, Halifax, Nova Scotia B3H 4J1, Canada

Possible determinants of the word-frequency effect (WFE), that is, the finding that low-frequency (LF) words are recognized more accurately than high-frequency (HF) words, are evaluated. Three studies examined the view that, since HF words have more meanings than LF words, it is less likely that the word sense tagged at time of presentation will be accessed at time of test and a correct response will be made. To ensure the same word sense was accessed at time of presentation and time of test in the case of both HF and LF words, sentence contexts used during presentation were combined with cuing procedures at time of test. Recognition performance improved, but the WFE was unaltered even when the identical sentence context was used during presentation and test. A fourth study considered Glanzer and Bowles' (1976) suggestion that associates of both HF and LF words tend to be HF words that are (1) likely to be activated and derivatively encoded during presentation and (2) likely to include a number of distractors from the recognition test sufficient to impair performance on HF words. Analysis of associative responses of subjects to HF and LF words, and the errors they made, support strongly such an interference-type interpretation of the WFE.

One of the most interesting paradoxes in memory research is the finding that if retention is tested by means of a recall test, high-frequency (HF) words are generally recalled more accurately than low-frequency (LF) words (e.g., Deese, 1960; Sumby, 1963), whereas on a recognition test, it is LF words that are most accurately designated as having been seen before (e.g., Allen & Garton, 1968; Glanzer & Bowles, 1976; Gorman, 1961; Underwood & Freund, 1970). It is puzzling that such different results are evident when retention of the same material is assessed via different tests. The existence of this paradox serves to raise the possibility that there is some fundamental and important difference between these two different procedures of memory assessment.

Numerous attempts have been made to provide an adequate account of the factors responsible for the word-frequency effect (WFE) (e.g., Eysenck, 1979; Glanzer & Bowles, 1976; Gregg, 1976). In these accounts, it is generally agreed that there are a variety of reasons to expect superior recall of HF words. For example, HF words are more meaningful; that is, in terms of Noble's (1952) definition, they have more immediately accessible associates. They are more available as responses (Underwood & Schulz, 1960). They can be accessed more readily in a variety of tasks, and organizational structures appear earlier in recall tasks

when HF words are used than when LF words are used (see Gregg, 1976).

Given the superior recall of HF words, it is surprising that these same words should fare so poorly on recognition tasks. Why should HF words be so poorly recognized? In a number of recent theoretical proposals, it has been suggested that at least some of the factors that likely facilitate the recall of HF words, such as the large number of different meanings and close associates they possess, may serve to impair their recognition.

The most systematic attempt to relate the recognition difficulties encountered with HF words to the large number of word meanings and associates they possess has been provided by Glanzer and Bowles (1976). They start with the assumption that when a word is presented in a memory experiment, some subset of the word's meanings, or "senses," are accessed and marked, or encoded. They assume that, during the course of accessing and encoding, there is a spread of activation leading to what they term "derivative" tagging of immediate associations of the to-be-remembered (TBR) word. During a recognition test, the process is repeated, and again, some subset of the total set of meanings of target and distractor words is accessed. It is the extent to which evidence of previous markings is apparent that determines whether a subject concludes that he has seen the word before. In the case of HF words, there are two difficulties. Because HF words have a greater number of possible word senses or encodings, it is less likely that the meanings or word senses accessed at the time of presentation and test will overlap, and adequate evidence of previous marking or tagging at time of presentation will be discerned. Further, it is likely that

This research was supported by a grant from the National Research and Engineering Council of Canada. The assistance of H. C. Ellis and the Department of Psychology of the University of New Mexico, Albuquerque, New Mexico, in providing facilities for the preparation of this manuscript is gratefully acknowledged.

in a substantial number of instances, distractor words will have been derivatively marked or tagged because they were activated as associates at the time of original encoding or presentation by some TBR word or another. Both these factors will serve to increase the difficulty of detecting and discerning the difference between TBR words and distractors.

In the case of LF TBR words, the above difficulties are much reduced. LF words have a smaller number of meanings than HF words have. Therefore, it is likely that the same word sense or meaning will be tagged at presentation and accessed at time of test. Moreover, it is contended that the associates activated at time of presentation to both HF and LF words will be HF words. Thus, although LF words will contribute to the difficulty of detecting HF words through derivative markings, they will be largely unaffected by derivative marking themselves.

The Glanzer and Bowles (1976) approach has substantial appeal. There is certainly evidence that HF words have more word senses than LF words have (e.g., Reder, Anderson, & Bjork, 1974; Schnorr & Atkinson, 1970). Further, their decision theory analysis of the pattern of forced-choice responses to all possible combinations of new and previously presented HF and LF words substantially strengthens the interpretation they offer. The present paper attempts to provide a further analysis of the view that the number of different word senses and derivative markings are major factors contributing to the WFE in recognition memory. The first three experiments are concerned with ascertaining the degree to which differences in the number of word senses, or meanings, contribute to the production of the WFE. The fourth experiment examines the character of free associations given to HF and LF TBR words, in an effort to evaluate assumptions about the derivative marking process. It attempts to determine the extent to which associative responses generated during presentation overlap with terms used on the recognition test. It attempts also to assess the degree to which subjects respond incorrectly on the recognition test to terms produced as free associates during the presentation sequence.

EXPERIMENTS 1 AND 2

Glanzer and Bowles (1976) maintain that it is the larger number of word senses that HF words possess and the likelihood that word senses accessed at time of test will differ from those marked at time of presentation that are largely responsible for the poor performance evident when HF words are presented on recognition tests. It follows, therefore, that if steps are taken to ensure that the same word sense encoded during presentation is accessed at time of test, the typically observed poorer performance on HF words should no longer be evident. In Experiments 1 and 2,

sentence contexts were used to ensure that a particular word sense of a TBR word was encoded at time of presentation. To ensure that the same word sense was accessed at time of test, appropriate semantic cues were provided on the recognition test. Experiment 2 differed from Experiment 1 only in that there was a 30-day separation between presentation and test. The expectation was that if access to the appropriate word sense was as important as we assumed, then provision of semantic cues at time of test should be more effective after a long delay, and under circumstances in which a relatively low retention level would be anticipated, than under conditions of immediate test.

Method

Design. There were three conditions in both experiments. The words-alone (WA) condition was included to demonstrate that the customary superiority of LF words would be evident with the materials and general procedures employed. Capitalized and underlined TBR words were presented singly, and retention was tested by administering a standard forced-choice recognition test on which each TBR word was paired with two distractors from the same frequency range. The sentence-context-with-cues (SWC) condition attempted to deal with the large number of meanings HF words possess and the likelihood that word senses at test overlap little with those tagged at presentation. Sentence contexts provided at time of presentation focused attention on a single dominant sense of each of the capitalized and underlined HL and LF words. To ensure that the appropriate word sense was accessed at time of test, each of the words on the recognition test was accompanied by a short phrase that directed attention to a specific meaning or sense of that word. In the case of TBR words, the word sense conveyed by the phrase was the same as that generated by the sentence context used at input. If, for example, the sentence context used for the TBR word "circle" was "The teacher instructed the students to use a compass to draw a circle with a radius of five inches," the word "circle" would appear on the retention test with two distractors and accompanying phrases, as shown below:

AVENUE	-----	a wide street
OPINION	-----	a belief or view
CIRCLE	-----	a geometrical form

A sentence-context-alone (SA) condition was included for control purposes. In this condition, sentence contexts were used at time of input, but no cue phrases accompanied words on the recognition test. By comparing this condition with the SWC condition, in which semantic cues were available at time of test, the effectiveness of providing semantic cues at time of test could be determined.

Materials. Two separate presentation lists (A and B) were constructed. Each list was composed of 40 LF and 40 HF words. LF words were considered to be those with a frequency of 1-5 times per million words of printed English according to the Thorndike-Lorge (1944) word count, and HF words were selected from those words occurring with a frequency of 49 or more times per million words. Differences between HF and LF words in terms of imagery, concreteness, and meaningfulness were minimized. According to the Paivio, Yuille, and Madigan (1968) norms, the mean value of HF and LF words on List A were, respectively, 4.92 and 4.97 for imagery, 4.62 and 5.06 for concreteness, and 5.95 and 5.70 for meaningfulness. The corresponding ratings for List B were 5.21 and 5.18 for imagery, 5.35 and 5.58 for concreteness, and 6.07 and 5.94 for meaningfulness. Three different random presentation orders of inter-

mixed HF and LF words were generated for both Lists A and B. Two different forms of a three-alternative recognition test were produced by pairing HF and LF words from Lists A and B together and adding another 40 HF and 40 LF words from the appropriate frequency range. Any set of three alternatives on the recognition test was always composed of words from the same frequency range. Three different random orders of the 80 HF and LF alternative sets were generated for each of the two different forms of the recognition test.

Sentence contexts emphasized a specific dominant meaning or sense of a TBR word. Two judges independently assessed each sentence context, and unless both judges were in accord, the sentence was rewritten so as to assume an acceptable form. A similar procedure was followed in forming the phrases to be used as cues on the recognition test in the SWC condition. The phrases were assessed also by two judges, and unless both agreed that a given phrase conveyed a dominant word sense (and in the case of the TBR word, the appropriate word sense), it was rejected and rewritten. Each TBR word was printed (capitalized and underlined) on a separate sheet of paper, alone or in the appropriate sentence context. These pages were then combined into appropriately ordered booklets for presenting words in the different conditions. To eliminate primacy and recency effects, the four pages of the booklet immediately prior to the first TBR word and the four pages following the last TBR word each contained a filler word of intermediate frequency. In the case of the SWC and SA conditions, filler words were in sentence contexts. A booklet format was used also for the recognition test. Each set of three alternative responses on a given page was separated from preceding and following pairs by dotted and horizontal lines. The three different presentation orders of Lists A and B were combined with three different orderings of the two different forms of the recognition test to produce 36 different presentation/test booklet combinations.

Subjects and Procedure. The study was carried out during the regular lecture periods of two undergraduate psychology classes. Both classes were large enough to provide the 36 subjects required in each of the three conditions used. The toss of a coin determined which class was to receive the immediate and which the delayed test. Prior to the arrival of students, the necessary presentation and test materials were placed on desks in each row in random order, with the restrictions that no condition could be repeated in a row until all conditions had been equally represented and that identical conditions could not appear on adjacent desks. Subjects were required to fill the room row by row, so as to ensure that approximately equal numbers would be assigned to all conditions.

After being seated, subjects were informed that the objective of the experiment was to assess how well they could remember words and that the large booklet in front of them (which they were instructed not to open) contained a large number of capitalized and underlined words that they would have to try to remember. They were informed that each page of the booklet would contain a separate word and that they would have 5 sec to examine the word on each page. They were then requested to read the instructions on the cover page of the presentation booklet. For subjects in the WA condition, in which words appeared alone, the instructions simply repeated those previously given; that is, each word was to be read carefully, and they were to try to remember it. In the SWC and SA conditions, in which sentence contexts were used, subjects were told that the capitalized and underlined words would appear in sentences and that it was extremely important that they read the sentence carefully and concentrate on remembering the capitalized and underlined word whose meaning was conveyed by the sentence context.

After all subjects had read the instructions, it was stressed that they were to turn the pages of the booklet one at a time, and only when the instruction to do so was given over the

loudspeaker. The instruction to turn the page was then given. A Hunter Klockcounter was used to time the subsequent 5-sec time intervals, and two assistants moved continuously about the class to be sure that page-turning instructions were followed. When inspection of the last page containing a word was completed, subjects were asked to turn to the page immediately following and print their name. This procedure provided a record of the presentation condition to which the subject had been assigned and allowed the experimenter to determine whether the subject had maintained the appropriate position in the booklet through the presentation sequence. After writing their names, subjects were requested to estimate the proportion of words that they thought the class as a whole would remember.

Subjects in Experiment 2, in which the recognition test was to be delayed, were told that the objective of the study was to see how closely estimates of the ability of college freshmen to retain material corresponded with their true ability and that estimates of what the class would remember would be compared with the actual performance of another class that would be given a retention test. The presentation booklets were then collected, and the lecture commenced.

Subjects in Experiment 1 were given a retention test immediately following their attempt to estimate the proportion of words that would be retained. They were asked to pick up test booklets beside their seats, read the instructions on the front cover carefully, and begin the test. The instructions informed them that the test would consist of a large number of sets of three words each and provided an illustrative example of how individual word sets would be separated from each other by dotted lines. Subjects were informed that one word in each set was from the presentation list. They were requested to carefully select and to circle the one word from each set that they thought most likely to have appeared on the presentation list. They were told to guess if in doubt, but not to move on to the next set of words until they had circled one word in the preceding set. The recognition test was unpaired.

In the SWC condition, subjects were instructed that the short phrases accompanying each word were intended to provide information to ensure that the appropriate sense or meaning of each word was considered in evaluating words on the recognition test. They were told to carefully consider the specified meanings of the words in determining which word to circle.

Subjects in Experiment 2 were given a delayed recognition test 30 days after the initial presentation of words. The test was given during a regular lecture period on a day on which high attendance was expected. Each subject obtained a recognition test booklet with his or her name on it from the experimenter. Codes on the presentation booklets signed earlier by each subject allowed appropriate assignment of different forms of the recognition test to participants. The procedures used in completing the test were identical to those described for subjects who received the immediate recognition test. After completion of the test, an explanation of the reasoning behind the procedures was provided and the objective of the study was outlined to all subjects.

In both Experiments 1 and 2, there were more subjects in the classes than required. This problem was dealt with by first examining the presentation booklets of subjects who completed the recognition test and deleting the data of any subject who failed to finish on the appropriate page of the presentation booklet. There were three such subjects in Experiment 1 and two in Experiment 2. Before any data analysis was undertaken, a table of random numbers was used to remove subjects required by the design. The data of 8 and 13 surplus subjects were discarded in Experiments 1 and 2, respectively.

Results and Discussion

The mean number of HF and LF words correctly

Table 1
Mean Number of Correct Responses (of 40 Possible)
on High- and Low-Frequency Words

Frequency	Experiment 1			Experiment 2		
	WA	SA	SWC	WA	SA	SWC
High	29.50	26.06	30.44	16.31	14.11	15.83
Low	34.58	32.78	35.39	18.33	17.53	17.75

Note—WA = words alone, SA = sentences alone, and SWC = sentences with cues.

recognized in Experiments 1 and 2 are shown in Table 1. The protracted delay between presentation and test impaired retention very substantially in Experiment 2. On average, 78.65% of the TBR words presented on the immediate recognition test given in Experiment 1 were correctly recognized, but only 41.61% of the words were correctly recognized after the delay imposed in Experiment 2. An analysis of variance performed on the number of correct responses attested to the reliability of the difference [$F(1,210) = 921.96$, $MSe = 25.71$, $p < .001$]. Although the retention levels differed greatly and the effects of the various experimental procedures appeared less pronounced in Experiment 2 than in Experiment 1, the overall pattern of the data was similar in the two experiments.

As would be expected from the literature, in Condition WA, in which HF and LF words were presented alone, LF words were more accurately recognized. Of substantially greater interest is whether the use of sentences to impose a specific semantic encoding at time of presentation and phrases to ensure that the appropriate semantic sense would be accessed at time of test reduced or eliminated the WFE. Table 1 shows that these procedures had little effect on the WFE. The magnitude of the difference between HF words and LF words in Conditions SWC and SA remained very similar to that evident in the WA condition in both experiments. An analysis of variance confirmed that LF words were more accurately recognized than HF words [$F(1,210) = 142.02$, $MSe = 12.28$, $p < .001$] and that there was no interaction between the WFE and conditions [$F(2,210) = 2.44$, $MSe = 12.28$, $p > .05$] or among the WFE, conditions, and experiments [$F(2,210) < 1$].

Contrary to expectations, providing semantic cues to ensure that the appropriate word sense was accessed on the recognition test was not more facilitative after a long delay. The interaction between conditions and delay was not significant [$F(2,210) = 2.25$, $MSe = 25.71$, $p > .05$]. In fact, small differences that were apparent were in the opposite direction.

Differences among the conditions reflecting the effect of providing sentence contexts and semantic cues at time of test were evident [$F(2,210) = 8.67$, $MSe = 25.71$, $p < .01$]. A Duncan multiple-range test was applied to compare individual conditions. The SWC condition differed from the SA and WA conditions at the 5%

level of significance, but the latter two conditions did not differ from each other. This pattern was anticipated. Any advantage accruing from the use of sentence contexts at time of presentation in the SA condition must be offset by the appreciable amount of additional information that has to be dealt with at time of input. Moreover, according to encoding-specificity theory (e.g., Thomson, 1972; Thomson & Tulving, 1970), the presence of a sentence context at time of presentation and its absence at time of test should also serve to reduce recognition accuracy. What is important is that the disadvantage associated with the imposition of a sentence context should be diminished by the provision of semantic cues at time of test. This was the case, and performance in Condition SWC, in which such cues were available, was superior to that observed in Condition SA.

The only interaction involving the WFE that did prove reliable involved experiments [$F(1,210) = 21.54$, $MSe = 12.28$, $p < .01$], indicating that the WFE was less pronounced in Experiment 2 than in Experiment 1. In previous experiments in which testing has been delayed, interactions between WFE and the duration of the delay have not proved reliable (see Gregg, 1976). Protracted speculation about possible reasons for the interaction in the present study appear unwarranted in view of the low retention levels evident in Experiment 2. The convergence of performance on HF and LF words was most likely induced by the low asymptotic performance levels approached as a result of the long delay separating presentation and test in Experiment 2.

Overall, Experiments 1 and 2 lend no support to the contention that the number of different meanings possessed by HF words is a major factor contributing to the WFE. When sentence contexts were used to ensure that a particular sense of a word was encoded at time of presentation and semantic cues were used to ensure that the same word sense was accessed at time of test, there was an improvement in recognition memory, but no change in the WFE under either immediate or delayed test conditions.

EXPERIMENT 3

Experiment 3 focused on the possibility that the manipulations used in Experiments 1 and 2 were too coarse; that is, although the appropriate general semantic senses of TBR words were conveyed at time of test, subtle gradations of a specific meaning or word sense were not conveyed suitably at time of test by the short phrases accompanying each word. Encoding-specificity theorists in particular might argue that, had a more precise encapsulation of the semantic sense encoded at input been conveyed at time of test, differences between HF and LF words might have been eliminated. Relevant to this point is the study by Fisher and Craik (1980), which found that complex sentence encodings were

more effective in facilitating recognition memory only if the test conditions were such that the original encoding context could be reintegrated. Experiment 3 constitutes an attempt to ensure that the encoding procedures employed at time of presentation coincide exactly with those employed at time of test.

Method

Design. There were two conditions. In the sentence-context-with-sentence-cue (SWSC) condition, an effort was made to ensure that the word sense accessed at time of test coincided more closely with that at time of presentation than was possible in Experiments 1 and 2, by including as part of the recognition test the same sentence context that was used at time of presentation. In the sentence-context-with-no cues (SWNC) condition, sentence contexts were provided at time of presentation, but TBR words and distractors appeared alone on the recognition test. By comparing the SWSC and the SWNC conditions, it was possible to gauge the effectiveness of presenting the same sentence context at time of presentation and time of test.

Materials. Presentation booklets containing words from Lists A and B in sentence contexts that were used in Experiments 1 and 2 were used in Experiment 3. New recognition test booklets were constructed. The rationale behind the procedures used in constructing the recognition test can be best explained by considering first the formation of a test item from the SWSC condition, in which sentence contexts constituted part of the recognition test. If, for example, the TBR word was an LF word from List A, then one of the four alternatives was the TBR word, capitalized and underlined, in the sentence context used at the time of presentation. A second alternative was an LF word drawn from List B, to which the subject had not been exposed, capitalized and underlined, in its sentence context. Two additional alternatives were constructed by removing the capitalized and underlined words from the two sentence contexts described and replacing them with words as similar in meaning and length as possible that would fit equally well into these sentence contexts.

An example of a typical recognition trial in the SWSC condition is shown in Table 2. The word "suds" is the LF word from List A and the word "debris" is an LF word from List B. The two remaining alternatives were formed by substituting the word "foam" in the sentence context used for the word "suds," and the word "rubble" in the sentence context used for the word "debris." The substitution procedure was necessary to take into consideration the possibility that subjects might not be able to recognize a TBR word but might remember having seen the sentence context. It was necessary, therefore, to include that identical sentence on the recognition test with an acceptable substitution, in order to deal with this possibility. The procedure of pairing LF and HF words from List A with corresponding HF and LF words from List B and then producing two additional alternatives by replacing TBR words with other words similar in meaning was followed for all words in Lists A and B. Two different judges assessed the acceptability of substituted words, and unless both agreed that the word substituted was suitably similar in meaning and that it fit equally well in the sentence context, the word was discarded and a replacement was found. After this procedure was completed, and additional re-pairing of alternatives was undertaken to produce two different forms of the recognition test.

Table 2
Sample Recognition Test Item From Experiment 3

FOAM from the bathtub overflowed onto the floor.
SUDS from the bathtub overflowed onto the floor.
There was a large pile of RUBBLE behind the gravel pit.
There was a large pile of DEBRIS behind the gravel pit.

Booklets with two different random orders of each of the two different forms of the 80-item forced-choice test were constructed. Recognition booklets for the SWNC condition, in which no sentence contexts were used at time of test, were identical to those used in the SWSC condition, except that sentence contexts were deleted.

Subjects and Procedures. Forty-eight students in an upper level undergraduate course served as subjects. Half the subjects received words in sentence contexts from List A and half, from List B. The two different forms and orders of the recognition test were assigned with equal frequency to subjects receiving List A and those receiving List B. In general, the procedures followed were identical to those followed in Experiments 1 and 2. Pages were turned in presentation booklets in accordance with instructions from the experimenter, and subjects were allowed 5 sec to examine each TBR word in its sentence context. It was emphasized that it was the underlined and capitalized words in each sentence that were to be remembered and that sentence contexts were intended to make clear the meaning of the TBR word.

The recognition test was administered immediately after completion of the presentation sequence and was unpaced. Instructions stressed the importance of examining all four capitalized and underlined words composing each set of alternatives before making a choice. Additional instructions to subjects in the SWSC condition emphasized the need to read all sentence contexts containing capitalized and underlined target words carefully before making a choice.

Results and Discussion

On average, subjects in the SWSC condition recognized 33.21 LF words and 29.00 HF words. In the SWNC condition, an average of 30.58 LF and 24.46 HF words were recognized. Performance was higher for subjects in the SWSC condition, in which sentence contexts were provided during presentation and test, than in the SWNC condition, in which sentence contexts were available only during the presentation sequence. Subjects in the SWSC condition recognized 77.76% of the TBR words, and subjects in the SWNC condition recognized only 68.80%. An analysis of variance performed on the number of correct responses shows this difference to be reliable [$F(1,46) = 4.85$, $MSe = 63.64$, $p < .05$] and the WFE to be reliable [$F(1,46) = 54.15$, $MSe = 11.83$, $p < .001$].

The issue of particular interest is whether provision of the input encoding context at time of test had any effect on the WFE. The magnitude of the WFE was slightly smaller in the SWSC condition, but the interaction between the WFE and the presence or absence of context at time of test did not approach significance [$F(1,46) = 1.86$, $MSe = 11.83$, $p > .05$]. The pattern of data was thus very similar to that observed in the preceding experiments. Providing the sentence context on the recognition test facilitates recognition memory, but it does not appreciably influence the WFE. Viewed collectively, the data from the three experiments described offer little support for the view that the WFE is closely related to the greater number of different word senses HF words possess and the difficulty subjects have in accessing or discerning that a given word sense was tagged at time of input when such words are presented. In the experiments reported, the information provided on the recognition test ranged from the appro-

priate general word sense (Experiments 1 and 2 to the reintroduction of the exact context provided at time of input (Experiment 3), but no effect on the WFE was apparent.

EXPERIMENT 4

On the basis of Experiments 1, 2, and 3, there is little evidence to suggest that differences in the number of different meanings or word senses contribute significantly to the WFE. What is responsible for the effect? One approach that remains to be examined concerns the possibility that a substantial amount of indirect marking or tagging of associates of TBR words occurs during the course of presentation. Underwood and Freund (1970) have speculated about this possibility, and it constitutes a major part of the analysis of the WFE offered by Glanzer and Bowles (1976). In their analysis of the WFE, Glanzer and Bowles propose that HF distractors on the recognition test are more likely to be indirectly marked or tagged during the presentation period than are LF distractors and that such indirect tagging must make the difficulty subjects encounter in differentiating between TBR words and their distractors much greater in the case of HF words than in the case of LF words. The assumption underlying this argument is that when a word is presented for study, not only are some of the senses or meanings of that particular word tagged, but in addition, there is an automatic spread of activation to other associated words and their semantic senses. To justify their argument, Glanzer and Bowles point out that the likelihood that a word will occur as an associative response increases with its frequency (Howes, 1957) and that there is evidence (Anisfeld & Knapp, 1968; Fillenbaum, 1969; Underwood, 1965) that subjects respond to implicit associations on recognition tests as they would to directly tagged target words.

Experiment 4 sought to examine the character of associative responses and the degree to which such responses could contribute to the WFE. Subjects were required to produce free associate responses to HF and LF TBR words before taking the recognition test. The protocols produced were then examined to determine the frequency with which TBR words and distractors used on the recognition test occurred as free associates. Analyses were also carried out to determine the extent to which errors made on the recognition test involved distractors and TBR words produced as free associates. Such a procedure can provide only a crude approximation of the spread of associative activation assumed to underlie the encoding phase of recognition memory. Moreover, the pattern of errors may not precisely mirror that which would be evident if the process occurred implicitly. Nevertheless, it would appear to provide the most direct means of externalizing the hypothesized internal associative process and assessing the degree to which it may contribute to the WFE.

Method

Design. There were two conditions. In both conditions, subjects were informed that a large number of words would be presented and that there would be a retention test to see how many of these words they could remember. In the free association (FA) condition, they were required to write on a line opposite to each word in the presentation booklet used, three free associates to the word. In the repetition (REP) condition, they were required to copy each word three times. This condition was designed to be used as a control to determine whether the procedure of requiring three free associates altered in any way the magnitude of the WFE that would normally be evident with a standard repetition procedure.

Materials. Preliminary experimentation indicated that having subjects provide free associations led to a very high level of performance. To avoid ceiling effects, words from Lists A and B were combined. The 36 LF and 36 HF words on each list were combined to produce a single 144-item list. Two different presentation orders of the 144 HF and LF words were generated, and booklets were constructed in which each word was listed on the left side of a page, with a line on the right to allow either free associations or repetitions of the words to be written. To avoid primacy and recency effects, four words of intermediate frequency were positioned at the beginning and end of each presentation booklet. Recognition test booklets were formed in the same fashion as in Experiments 1 and 2. Each target word was combined with two alternatives from the same frequency range. There were two combinations of TBR words and distractors.

Subjects and Procedures. Twenty-four undergraduates were recruited and paid for participating in the study. Half of the subjects were randomly assigned to the FA condition and the other half, to the REP condition. Subjects were tested independently. They were informed that their memory for words appearing on the left-hand side of the page of the presentation booklet would be tested. They were asked to proceed through the booklet word by word, either writing the word on the left three times or providing three free associates, depending on the condition to which they were assigned. Subjects in the FA condition were asked to avoid abbreviations and proper names and to minimize repetitions. After completion of the booklet, an appointment was set up for the next day at the same time and the forced-choice recognition test was administered. Instructions used on the test were identical to those used in Experiments 1 and 2, and the test was unpaired.

Results and Discussion

Requiring subjects to provide three free associates to each word resulted in a much higher overall performance than having them write the word three times. After 24 h, subjects in the FA condition recognized 63.96 words, whereas subjects in the REP condition retained only 45.38 of the words presented. An analysis of variance performed on the number of correct responses established the reliability of this difference [$F(1,22) = 103.42$, $MSe = 40.07$, $p < .001$].

The reason that the difference between the FA and REP condition was so extreme is not clear. It is possible that the FA procedure simply forced a more elaborative encoding on subjects. It is also possible that it is the self-generation part of the process that is important (see Slamecka & Graf, 1978).

Of more immediate interest is the question of whether the WFE was evident in Experiment 4 and whether the magnitude of the effect differed in the REP

and FA conditions. Subjects in the FA condition recognized correctly 68.67 LF words and 59.25 HF words. Subjects in the REP condition recognized 52.33 LF words and 38.42 HF words correctly. Analysis of variance confirmed the presence of a reliable WFE [$F(1,11) = 88.29$, $MSe = 18.50$, $p < .01$] and that there was no interaction between the WFE and the two different conditions employed [$F(1,22) = 3.28$, $MSe = 18.50$, $p > .05$].

Since the FA procedure did not produce a WFE that differed reliably in magnitude from that which would be typically encountered under a standard repetition condition, it is possible to proceed to the main objective of the study, which was to provide information about the relationship between the free associates generated in the FA condition and the distractors and TBR words on the recognition test.

The protocols of all subjects in the FA condition were examined, and all the different words generated that occurred on the recognition test as either distractors or targets were tabulated for each subject. The number of errors made on distractors and TBR words produced as free associates were also recorded. These data are shown in Table 3.

The basic issues of interest are (1) whether HF distractors and/or TBR words occur more commonly as free associates than LF distractors and TBR words do and (2) whether a substantial number of errors are made in mistaking distractors produced as free associates for TBR words, or, in the case of TBR words generated as distractors, whether these words are confused with distractors and not identified correctly on the recognition test.

With respect to the first issue, it is evident from the table that, overall, HF distractors were more commonly produced as free associates than LF distractors were. On average, 20.94% of the 144 HF distractors were generated as free associates, whereas only 1.10% of the LF distractors were produced. Similarly, on average, 15.28% of the 72 HF TBR words occurred as free associates, whereas only 2.31% of the LF TBR words occurred as free associates. These differences were reliable. Separate

analyses of variance were performed on distractors and TBR words that were produced as free associates. A significantly greater number of HF distractors than of LF distractors were produced as free associates [$F(1,11) = 564.75$, $MSe = 4.34$, $p < .001$], and a significantly greater number of HF TBR words than of LF TBR words were generated as free associates [$F(1,11) = 171.93$, $MSe = 1.52$, $p < .001$].

A more detailed scrutiny of the data summarized in Table 3 shows that, in general, more distractors were produced as free associates to HF TBR words [$F(1,11) = 18.52$, $MSe = 7.75$, $p < .01$]. A similar tendency for more TBR words to be produced as free associates to other HF TBR words than to LF TBR words fell short of significance [$F(1,11) = 4.08$, $MSe = 3.45$, $p > .05$]. HF distractors were given more commonly as free associates to HF TBR words, and LF distractors (although few in number) occurred more commonly as free associates to LF TBR words [$F(1,11) = 36.24$, $MSe = 6.83$, $p < .01$]. HF TBR words were given more commonly as free associates to other HF TBR terms, and LF TBR words, which were also few in number, were given more frequently to other LF TBR terms [$F(1,11) = 8.21$, $MSe = 4.48$, $p < .01$]. No qualification of the main finding that HF distractors and TBR words appeared much more frequently as free associates than LF terms did is required as a result of the more detailed scrutiny of the data provided above. Therefore, the second question, concerning the degree to which subjects make errors by confusing distractors with TBR terms and TBR terms with distractors, can be addressed.

Table 3 shows the average number of errors made on distractors and TBR words produced as free associates. A very considerable number of HF distractors generated as free associates were mistaken for TBR words. On average, 20.99% of the HF distractors generated as free associates were mistakenly designated as TBR words. With respect to HF TBR words generated as free associates, an average of 9.82% of these words were mistaken for distractor items and not circled on the recognition test. Errors involving LF distractors and TBR terms generated as free associates were made also. For example, 31.65% of LF distractors generated were mistaken for TBR words, but the number of LF distractors generated was so small that the total number of errors produced amounts, on average, to a mere .50 of a word, and there were 144 words that served as LF distractors. The point to be made is that the number of LF terms generated as free associates was so small that their overall contribution to the error pattern was minimal.

The largest number of distractors and TBR words produced as free associates were HF words, and it is with respect to these words that most of the errors occurred. Were they sufficient in number to constitute a major determinant of the WFE? An affirmative answer is suggested by the observation that 58% of the HF errors made involved the 19.05% of the 144 HF dis-

Table 3
Mean Number of Distractors and TBR Terms on the Recognition Test Generated as Free Associates, and the Mean Number of Errors Involving These Terms on the Recognition Test

F	Distractors Generated		Distractor Errors		TBR Words Generated		TBR Word Errors	
	H	L	H	L	H	L	H	L
H	19.08	.25	4.00	.00	6.92	.50	.50	.17
L	11.08	1.33	2.33	.50	4.08	1.16	.58	.08
Total	30.16	1.58	6.33	.50	11.00	1.66	1.08	.25

Note—F = frequency of words to which free associates were given; H = high, L = low. Each subject produced 432 free associates. There were 144 distractors and 72 TBR words at both the high- and low-frequency levels on the recognition test.

tractors and 72 HF TBR words produced as free associates. Statistical analysis confirms that the average 7.42 errors made to HF distractors and TBR terms that were generated as free associates differs reliably from the average 5.33 errors made to HF distractor and TBR terms not produced as free associates [$F(1,11) = 5.30$, $MSe = .23$, $p < .05$, after a square-root transformation].

Obviously, not all HF errors produced can be related to HF terms generated as free associates. An average of 5.33 errors cannot be directly related to free associates generated. Some of these residual errors may reflect deficiencies in our efforts to externalize assumed internal associative processes. For example, it is possible that in some instances, a TBR word will elicit an abstract representation that cannot be immediately instantiated as a word when free associates are required, but which may later resonate with a distractor on the recognition test and lead to an incorrect response. The source of such responses must escape notice with the procedures employed. It is also possible that residual errors reflect, in part, the contribution of another factor (or factors) to the WFE. These caveats notwithstanding, the overall pattern of data produced by the operations employed corresponds closely to that which would be expected if derivative associate marking were an important determinant of the WFE. More HF distractors and TBR words than LF distractors and TBR words were produced as free associates. Further, the majority of errors were associated with HF words produced as free associates, even though they constituted a relatively small component of the entire set of HF distractor and TBR words.

GENERAL DISCUSSION

The finding that recognition memory can be dramatically affected by contextual change (Light & Carter-Sobel, 1970) has led to a great deal of speculation about the possibility that it is not the word as a unitary whole that is encoded at time of input, but rather, a particular semantic aspect or interpretation of the word. This approach handles context effects nicely. Recognition failure may be anticipated if the contexts at input and test are different and evoke different semantic senses of the word. The view that word senses are encoded has become very widespread (e.g., Baddeley, 1976; Pellegrino & Salzberg, 1975). Moreover, the view that word senses are encoded is not confined exclusively to experiments in which there is a direct manipulation of context, but such encoding is also assumed to occur when words are presented alone. It follows that, just as recognition failure can occur in the situation in which the context is different at the time of presentation and time of test, so in the case of a single word will recognition failure occur if there is not sufficient overlap between the semantic sense or area of semantic space encoded at time of presentation and that accessed at time of test.

Under such circumstances, as Glanzer and Bowles

(1976) and Reder et al. (1974) point out, LF words should have a distinct advantage over HF words in recognition memory studies. LF words have fewer different senses or meanings than HF words have, and it is much more likely that the same sense will be accessed at presentation and at test than would be the case with HF words. A necessary consequence, explicitly recognized by Glanzer and Bowles, is that if steps are taken to restrict the number of senses sampled at time of input and to ensure that the same senses are activated at time of test, the difference in recognition memory normally evident between HF words, which have many senses, and LF words, which have few senses, should be markedly diminished. In Experiments 1, 2, and 3, steps were taken to ensure that the same word sense was accessed at time of presentation and time of test. No support was obtained for the position that accessing the appropriate word sense at the time of presentation and time of test constitutes a major determinant of the WFE.

The present data do not stand alone in suggesting that the number of word senses possessed by HF and LF words is not a crucial determinant of the WFE. Eysenck (1979) makes brief reference to a study in which he attempted to control the number of different meanings possessed by HF and LF words and still found evidence of the WFE. Also, relevant to the sense-accessing issue is an investigation of the encoding-specificity effect by Tulving and Watkins (1977), in which words were carefully selected so as to have a single meaning. With such words, the same semantic sense should be activated at presentation and at test, and recognition failure should not be evident. Nevertheless, Tulving and Watkins found that subjects recalled words that they could not recognize as words they had seen before. A final point is that if sense tagging were as crucial to the recognition process as has been assumed, it would be expected that increasing the number of different word senses tagged and providing cues to ensure that appropriate senses are accessed at time of test would dramatically improve recognition memory. The use of these procedures in a previous experiment (Earhard & Landry, 1976) resulted in a statistically reliable, but numerically small, improvement. Similarly, in the present study, the effect of providing context and cues so as to ensure that the appropriate word sense is activated reliably facilitates performance, but the magnitude of the effect, especially after a long delay, is extremely small compared with what would be expected if access to the appropriate word sense were a major problem in recognition memory.

Although little support for the view that the number of different word senses possessed by HF and LF words contributes substantially to the WFE, the results of Experiment 4 suggest that the implicit tagging of associates of HF words during presentation does likely play an important role in the production of the WFE. The majority of HF errors in the experiment were associated with distractors and TBR words generated as free associates.

How representative is Experiment 4 of the assumed implicit associative process? More specifically, does requiring that three free associates be given to every TBR word magnify the role of the associative marking process as a contributor to the WFE? It is certainly reasonable to assume that the automatic spreading-activation process assumed to underlie the marking of implicit associations will, under some circumstances, lead to the tagging of more than three free associates and, under other circumstances, to fewer than three. It seems unlikely, however, that the procedure of asking for three free associates in Experiment 4 magnifies the role of implicit associative marking. The magnitude of the WFE in Experiment 4 was no larger when three free associates were required to TBR words than when TBR words were simply written three times. Moreover, there is a considerable body of other data that mesh very well with the view that implicit tagging of HF associates during presentation plays a very important role in the production of the WFE.

Balota and Neely (1980) report significantly higher false alarm rates on HF than on LF lures. Similar evidence that subjects are more likely to falsely recognize HF words than LF words is provided by Glanzer and Bowles (1976) and McCormack and Swenson (1972). The findings of Underwood and Freund (1970) that HF words are more difficult to recognize than LF words only when HF words are used as distractors and that HF words with LF distractors are recognized more accurately than LF words with HF distractors would also be expected if the WFE stems largely from HF associative responses generated during presentation that are confused with HF distractors and TBR words at time of test. This position also makes sense of the data provided by Schulman (1976), which show that if words are sufficiently rare, the WFE is reversed and words judged more familiar are recognized better than words judged to be less familiar. The words employed by Schulman are so rare that the possibility of any distractor or TBR word being generated as an implicit associative response is negligible, and hence, a major determinant of the WFE is absent.

The WFE becomes less of a paradox if the above approach is taken. In fact, if the major determinant of the WFE is interference at time of test generated by HF associative responses that occur during the presentation process, the HF words may well prove to be more accurately recognized at time of test than LF words if the intrusion of associative responses can be prevented. The findings of Schulman (1976) and Underwood and Freund (1970) lend a measure of support to this contention. To put it another way, LF words may be more distinctive on a recognition test than HF words are, not because they are processed more elaborately, have a particular orthography, or have fewer meanings (see Eysenck, 1979), but because they are presented against a less confusing background

than LF words are. If the background confusion can be eliminated, then the superior performance of subjects on LF words in recognition memory should also be eliminated.

REFERENCES

- ALLEN, L. R., & GARTON, R. E. The influence of word knowledge on the word-frequency effect in recognition memory. *Psychonomic Science*, 1968, **10**, 401-402.
- ANISFELD, M., & KNAPP, M. Association, synonymy, and directionality in false recognition. *Journal of Experimental Psychology*, 1968, **77**, 171-179.
- BADDELEY, A. D. *The psychology of memory*. Basic Books: New York, 1976.
- BALOTA, D. A., & NEELY, J. H. Test expectancy and word-frequency effects in recall and recognition. *Journal of Experimental Psychology: Human Learning and Memory*, 1980, **6**, 576-587.
- DEESE, J. Frequency of usage and the number of words in free recall: The role of association. *Psychological Reports*, 1960, **7**, 337-344.
- EARHARD, B., & LANDRY, D. Dual vs. single encoding in recognition memory as a function of lag, delay, and availability of semantic information. *Memory & Cognition*, 1976, **4**, 609-615.
- EYSENCK, M. W. Depth, elaboration, and distinctiveness. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing in human memory*. Hillsdale, N.J.: Erlbaum, 1979.
- FILLENBAUM, S. Words as feature complexes: False recognition of antonyms and synonyms. *Journal of Experimental Psychology*, 1969, **82**, 400-402.
- FISHER, R. P., & CRAIK, F. I. M. The effects of elaboration on recognition memory. *Memory & Cognition*, 1980, **8**, 400-404.
- GLANZER, M., & BOWLES, N. Analysis of the word-frequency effect in recognition memory. *Journal of Experimental Psychology: Human Learning and Memory*, 1976, **2**, 21-31.
- GORMAN, A. M. Recognition memory for nouns as a function of abstractness and frequency. *Journal of Experimental Psychology*, 1961, **61**, 23-29.
- GREGG, V. Word frequency, recognition and recall. In J. Brown (Ed.), *Recall and recognition*. London: Wiley, 1976.
- HOWES, D. On the relation between the probability of a word as an association and general linguistic usage. *Journal of Abnormal and Social Psychology*, 1957, **54**, 75-85.
- LIGHT, L. L., & CARTER-SOBELL, L. Effects of changed semantic context on recognition memory. *Journal of Verbal Behavior and Verbal Learning*, 1970, **9**, 1-11.
- MCCORMACK, P. D., & SWENSON, A. L. Recognition memory for common and rare words. *Journal of Experimental Psychology*, 1972, **95**, 72-77.
- NOBLE, C. E. An analysis of meaning. *Psychological Review*, 1952, **59**, 421-430.
- PAIVIO, A., YUILLE, J. C., & MADIGAN, S. A. Concreteness, imagery, and meaningfulness for 925 nouns. *Journal of Experimental Psychology Monograph Supplement*, 1968, **76**(1, Pt. 2).
- PELLEGRINO, J. W., & SALZBERG, P. M. Encoding specificity in cued recall and context recognition. *Journal of Experimental Psychology: Human Learning and Memory*, 1975, **1**, 261-270.
- REDER, L. M., ANDERSON, J. R., & BJORK, R. A. A semantic interpretation of encoding specificity. *Journal of Experimental Psychology*, 1974, **102**, 648-656.
- SCHNORR, J. A., & ATKINSON, R. C. Study position and item differences in the short- and long-term retention of pairs associates learned by imagery. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 614-622.
- SCHULMAN, A. I. Memory for rare word previously rated familiarity. *Journal of Experimental Psychology: Human Learning and Memory*, 1976, **2**, 301-307.

- SLAMECKA, N. J., & GRAF, P. The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory*, 1978, 4, 592-604.
- SUMBY, W. M. Word frequency and serial position effect. *Journal of Verbal Learning and Verbal Behavior*, 1963, 1, 443-450.
- THOMSON, D. M. Context effect in recognition memory. *Journal of Verbal Learning and Verbal Behavior*, 1972, 11, 497-511.
- THOMSON, D. M., & TULVING, E. Associative encoding and retrieval: Weak and strong cues. *Journal of Experimental Psychology*, 1970, 86, 255-262.
- THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words*. New York: Bureau of Publications, Teachers College, 1944.
- TULVING, E., & WATKINS, O. C. Recognition failure of words with a single meaning. *Memory & Cognition*, 1977, 5, 513-522.
- UNDERWOOD, B. J. False recognition produced by implicit verbal responses. *Journal of Experimental Psychology*, 1965, 70, 122-129.
- UNDERWOOD, B. J., & FREUND, J. S. Word frequency and short-term recognition memory. *American Journal of Psychology*, 1970, 83, 343-351.
- UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning*. Philadelphia: Lippincott, 1960.

(Received for publication April 7, 1981;
revision accepted November 11, 1981.)