Journal of Experimental Psychology
1974, Vol. 102, No. 4, 648–656

A SEMANTIC INTERPRETATION OF ENCODING SPECIFICITY[1]

LYNNE M. REDER,[2] JOHN R. ANDERSON, AND ROBERT A. BJORK

University of Michigan

Two experiments are presented to clarify possible interpretations of the Encoding Specificity Principle of Tulving and Thomson. This principle states that a cue must have been studied with a word in order for the cue to be effective at testing. In the experiments reported here, recall and recognition of words were impaired by a change in the accompanying cues only if the to-be-remembered (TBR) words were of high frequency; low-frequency words did not support the Encoding Specificity Principle. The data suggest that both recall and recognition of a TBR word depend upon recognition of a specific interpretation of the word originally encoded, rather than its physical representation.

The considerable recent research on context effects in memory (e.g., Bobrow, 1970; Light & Carter-Sobell, 1970; Thomson, 1972; Thomson & Tulving, 1970; Tulving & Thomson, 1971; Winograd & Conn, 1971) has demonstrated that the ability to recall or recognize a word can be severely impaired by changes from study to test of the context in which the word is presented. Light and Carter-Sobell, for example, paired homographs with one adjective at the time of study (e.g., soda *cracker*) and later tested recognition of the homographs when paired with the same adjective or with an adjective that primed a different sense of the homograph (e.g., safe *cracker*). Recognition was significantly better when the same adjective was paired with the target word even though *S* knew that only the noun was to be judged for familiarity; thus, a word's context can greatly affect whether *S*s can access information stored with that word. This result is significant, since many models of memory (e.g., Anderson & Bower, 1972; Kintsch, 1970) assume that accessing a word's trace is an automatic process in tests of recognition memory.

Tulving and Thomson (1971) and Thomson (1972) have also found that recognition memory for to-be-remembered (TBR) words is impaired by changes in the cue word that is paired with a TBR word. Also, they found that adding or subtracting the cue word at the time of test impaired the recognition of TBR words. In contrast to the Light and Carter-Sobell (1970) procedure, however, Tulving and Thomson did not use transparent homographs, and they did not make a deliberate attempt to prime different meanings of the words. Thomson and Tulving (1970) also have found that cued recall of TBR words is lowered when the cue employed at the retrieval test is different from the encoding cue given at input, and that result holds even when the retrieval cue is a strong normative associate of the TBR word. Strongly associated, extralist cues (*strong cues*) enhance recall when no cues are presented with the TBR words at input (Bahrick, 1969; Thomson & Tulving, 1970), but are inferior probes compared to weak cues that were paired with TBR words at the time of input. It is possible in the Thomson and Tulving cued-recall situation that the *S* implicitly generates the TBR word to the strongly associated cue, but does not recognize the TBR word as being from the list. This interpretation is supported by the fact that *S*s can generate the TBR item in a free-association task to a strong cue, yet fail on a later test to recognize that

[1] This research was supported by the Air Force Office of Scientific Research under Contract F44620-72-C-0038 with the Human Performance Center, University of Michigan, and by a National Institute of Mental Health traineeship and a National Science Foundation graduate fellowship to the first author. We wish to thank Ed Martin for commenting on an earlier version of the manuscript.

[2] Requests for reprints should be sent to Lynne M. Reder, Human Performance Center, 330 Packard Road, Ann Arbor, Michigan 48104.

test to recognize that word as a TBR word (Tulving & Thomson, 1973).

Tulving and Thomson (1973) argue that only their Encoding Specificity Principle is consistent with these results. In defining this principle, they state that: "In its broadest form the principle asserts that only that can be retrieved that has been stored, and that how it can be retrieved depends on how it was stored [p. 359]." They assert that TBR words are stored as higher order episodic units and only elements of the *episodic* as opposed to *semantic* (see Tulving, 1972) higher order unit can access the TBR item. However, Light and Carter-Sobell (1970) reject the notion that their results could be due simply to "the effect of breaking up an encoded adjective-noun unit at recognition [p. 8]." They point out that words were recognized less frequently when a new adjective changed the sense of the noun than when the original adjective was simply deleted, and that recognition was better when the original adjective (e.g., soda *cracker*) was replaced by one that did not drastically alter the semantic interpretation of the target word (graham *cracker*) than when the new adjective did alter the semantic interpretation of the target word in a substantial way (safe *cracker*). Thus, Light and Carter-Sobell interpret their results as evidence that recognition of a word in a verbal learning experiment may involve recognition of the appropriate meaning of the word.

Work by Winograd and Conn (1971) supports the Light and Carter-Sobell (1970) interpretation. Words known to have multiple meanings were presented to *S*s with no explicit context, that is, no encoding cue. Prior to the main experiment, the relative frequencies of the various interpretations of each word were determined. Recognition for these words was then tested by presenting a word in 1 of 3 ways: (*a*) In a sentence that used a frequent meaning or interpretation of the word; (*b*) in a sentence that used an infrequent meaning of the word; or (*c*) by itself, that is, with no context or meaning imposed by *E*. Homographs tested in a context that utilized their high-frequency meanings and those tested without an imposed context were recognized about equally well, and both were recognized far more often than the homographs forced into an uncommon encoding or meaning. Winograd and Conn concluded that words with multiple meanings are encoded with respect to a specific meaning even though no cue is presented at input and that, in the absence of an experimentally provided context, the most frequent meaning of a word tends to be the one encoded.

It might appear that the Light and Carter-Sobell (1970) and the Winograd and Conn (1971) analyses do not extend to the Tulving and Thomson data because Tulving and Thomson did not use explicit homographs. However, many words in the language that are not explicit homographs still have multiple senses or interpretations. For instance, *water* does not have 2 obviously different meanings, but the sense of *water* when presented with *lake* is somewhat different than the sense of *water* when presented with *drink*. In fact, *Webster's Seventh New Collegiate Dictionary* (1965) lists 8 main senses of the noun *water* and 16 subsenses and 19 sub-subsenses Included are senses unfamiliar to most people and some obscure senses—e g., "a capital stock not representing assets of the issuing company and not backed by earning power [p. 1006]"—that are unrelated to any of the normal senses of the word. It is also certainly the case that there are still other interpretations of the word that are not in the dictionary but are known to some *S*s. Thus, many words not considered homographs allow multiple semantic interpretations, and a plausible explanation of the Tulving and Thomson data posits that a change in cue words can tap a different sense of the TBR word.

The experiments reported here were designed to test this semantic interpretation of encoding specificity. The basic idea is that words with few senses should be less vulnerable to changes in context in the

Tulving and Thomson paradigm than should words with many senses.

Although the idea behind this research is relatively straightforward, obtaining an adequate, objective measure of the number of senses associated with words is a nontrivial methodological problem. One cannot simply ask *S*s to report how many senses they have for particular words; it is not easy for *S*s to retrieve all the senses they have for a word (MacKay & Bever, 1967), nor is it easy to convince them that subtle distinctions like those associated with *water* are to be regarded as different senses. If it were easy to retrieve all the senses of a word, *S*s would not suffer the difficulties they do in experiments like those of Light and Carter-Sobell (1970).

For other reasons, number of dictionary meanings is also an inadequate measure. Many senses listed in a dictionary are not instantiated in most people's lexicon, while other colloquial and idiosyncratic senses are not in the dictionary. It is also difficult in practice to be consistent in terms of the cutoff point one uses to decide that a dictionary distinction is in fact substantial enough to denote separate senses of a word.

Based on these and other considerations, word frequency was chosen as the index of number of word senses. Words of very low frequency (less than 9/1,000,000 in Kučera & Francis, 1967) were contrasted with words of moderate to high frequency (more than 50/1,000,000). Generally, low-frequency words (e.g., hippopotamus, aspirin) tend to have only one sense. This phenomenon is probably attributable to there being little opportunity for multiple senses to differentiate and be maintained. It may also reflect the fact that words with one unique sense are appropriate to only a few situations and as a result have low frequency of occurrence. Therefore, word frequency is a fairly sensitive correlate of number of senses instantiated in the typical *S*. Schnorr and Atkinson (1970) found a strong positive correlation between word frequency and number of dictionary meanings, as did we for the words used in the experiments reported here, $r = .61$; $t\ (118) = 8.3$, $p < .001$.[3] Both word frequency and number of dictionary meanings are undoubtedly imperfect reflectors of number of senses, but frequency seems to provide a somewhat more satisfactory and objective basis on which to partition words into those having few or many senses. Later in this report frequency and number of dictionary meanings are compared in terms of their ability to predict memory performance.

Experiment I

Method

Design and procedure. Four lists of 30 TBR words, 15 high frequency and 15 low frequency, were presented by slide projector at a 3.5-sec. rate to *S*s for study and subsequent recall. Each TBR word was initially paired with a weakly associated cue; the TBR word was shown in uppercase letters directly below the cue word, which was shown in lowercase letters. After presentation of 30 pairs, 10 of the TBR words were tested in each of the 3 following conditions: (*a*) The *S* saw the original weakly associated input cue and was asked to recall the TBR word that appeared with it; (*b*) *S* saw a new cue strongly associated with a TBR word and was asked to use it to help him think of a TBR word from the last list; or (*c*) *S* received no cue for the word and was simply asked to try to recall the item from memory. Within each list, the 2 levels of word frequency were crossed with the 3 types of output conditions. There were 5 words in each Frequency × Output combination. Three different sets of recall sheets were used for each list so that, across *S*s, all words appeared in all output conditions. Each set consisted of 2 recall sheets, one for cued recall and one for the recall of words for which no probe or cue was given. The sheet for cued recall was presented before the free-recall sheet for half of the *S*s and after the free-recall for the other *S*s. The *S*s had as much time as they wanted for both uncued and cued recall. When uncued recall preceded cued recall, *S*s were asked to recall all the TBR words; when uncued recall followed cued recall, *S*s were told that they only needed to recall those words from the list that they had not recalled on the cued-recall sheet, but that they need not worry about duplications.

The word presentation order within a list and the order of the cue words on the cued-recall

[3] We had a group of *S*s rate all the words in our experiments for number of different meanings they could think of for each word. The *S*s were only able to generate a mean of 1.93 meanings per word whereas Webster's dictionary has 8.83 meanings. The number of generated meanings correlates .64 with frequency and .67 with number of dictionary meanings.

sheets were randomly determined but were held constant across *Ss*. Half of the 20 cues were asterisked to indicate that they were new cue words. The various recall sheets and the order of list presentation were counterbalanced over *Ss*. Two priming lists, which were not scored, preceded the 4 lists of interest. The priming lists, composed of medium-frequency words, were tested at recall with their original input cues. The purpose of the priming lists was to induce *Ss* to encode the TBR word with respect to the original cue presented; *Ss* were not informed that the priming lists were practice lists. Forty-eight paid *Ss* at the University of Michigan were tested in groups of 1–7.

Materials. All TBR words were nouns judged to be high in both meaningfulness and imagery. Items were selected from norms of frequency of words in print (Kučera & Francis, 1967); the high-frequency words came from the upper third of the rank ordering in frequency, and the low-frequency words came from the bottom third; words in the priming lists were taken from the middle range. (The high-frequency words all had A or AA Thorndike & Lorge, 1944, ratings, and the low-frequency words averaged 6.7/1,000,000 in the Thorndike and Lorge count.) Those nouns selected that were included in Paivio, Yuille, and Madigan's (1968) rating of meaningfulness and imagery of nouns received scores that confirmed the subjective criterion for high imagery and concreteness: They all were above 6 on the two 7-point scales. Thomson and Tulving (1970) constructed their lists of TBR words by selecting response words from free-association norms in which each TBR word occurred twice in the norms: once as a high-frequency response to a *strong cue* stimulus word and once as a low-frequency response to a different *weak cue* stimulus word. Association norms could not be used to select the cues for this experiment due to the difficulty of finding norms that contain low-frequency nouns as responses to other words, particularly as strong associates.

Because of these inherent difficulties, strong and weak cues were generated according to intuitive criteria. That is, the strong cues seemed likely to elicit the TBR words frequently, the weak cues infrequently. The fact that the cues were constructed in an informal manner does not seem a serious problem in methodology. The basic requirement is simply that the strong cues for low-frequency words are not more effective than strong cues for high-frequency words. This requirement is important because the predicted results (namely, that recall of low-frequency words with their new, strong cues will be superior to recall of high-frequency words with their new, strong cues) might otherwise be an artifact of prior strength. Five colleagues, naive with respect to the study, gave free associations to each of the strong cues used in the experiment. High-frequency TBR words were generated as free associates to the strong cues 24% of the time; low-frequency TBR words were generated only 12% of the time.

TABLE 1

PROPORTIONS OF HIGH- AND LOW-FREQUENCY TO-BE-REMEMBERED WORDS RECALLED ON TESTS OF CUED AND FREE RECALL

Word frequency	Cued recall		Free recall
	Strong extralist cue	Weak within-lists cue	
Cued recall before free recall			
High	.31	.69	.04
Low	.64	.72	.05
Cued recall after free recall			
High	.36	.63	.12
Low	.59	.69	.14

The other cue-construction requirement used by Thomson and Tulving (1970) is that the strong and weak cues for a given TBR word should not be associatively related. This requirement was impossible to maintain with low-frequency words. For 2 words to be both associatively related to a third word and yet not be associatively related themselves tends to require that the third word have more than one interpretation, and low-frequency words often do not. The only requirement in the current experiment was that no 2 cues to a given word be synonyms.

Results and Discussion

In Table 1, recall probabilities are shown as a function of type of test, order of tests, and TBR word frequency. The statistical analyses reported in this section were carried out on arc sine transformations of the proportions in Table 1.

Cued recall. Since performance on tests of cued recall was not sensitive to testing order, F (1, 46) = 1.2, the other statistical analyses of the cued-recall data were carried out on the pooled data. The result of greatest interest in this experiment, that is, the frequency by cue (strong vs. weak) interaction, was highly significant, F (1, 46) = 18.3, $p < .001$, with performance in the high-frequency strong-cue condition much worse than performance in the other conditions. For high-frequency words, strong extralist associates were much less effective as recall cues than were weak intralist associates, F (1, 46) = 77.44, $p < .001$. In the case of low-frequency words, there is a smaller but still signifi-

cant advantage in the same direction, F (1, 46) = 7.56, $p < .01$. The latter result is not too surprising since low-frequency words are sometimes polysemous, and in such cases a change of cue would have adverse effects.

In order to rule out the possibility that the foregoing results might be attributable, in part, to item selection (cf. Clark, 1973), the data were also analyzed by collapsing over *S*s and treating words as the random effect. The Critical Frequency × Cue interaction was still highly significant, F (1, 118) = 13.51, $p < .001$. Even with both significant *F*s, however, the results may still be colored by *S*–item selection. Clark prescribes the use of a *min F′* statistic, $(F_1F_2)/(F_1 + F_2)$, a conservative test, the significance of which guarantees generalizability of a result over *S*s and items. The *min F′* statistic is significant for both the Frequency × Cue interaction, F' (1, 149) = 7.77, $p < .01$, and for the effect of weak vs. strong cues for high-frequency words, F' (1, 135) = 37.91, $p < .001$. On the basis of this more conservative test, however, the effect of weak vs. strong cues was not significant for low-frequency words, F' (1, 142) = 3.45.

Free recall. The results of the free-recall test in Experiment I are relatively straightforward. Overall, *S*s were a great deal less likely to free recall the 10 TBR words not tested on the cued-recall sheets than they were to recall TBR words in response to either type of cue. In contrast to the cued-recall performance, there was a distinct effect of test order, F (1, 46) = 41.9, $p < .001$. The test of cued recall appears to have provided substantial interference with subsequent free recall of TBR words not tested during the cued recall, whereas cued recall was independent of whether there was or was not a prior free recall.

Experiment II

Our interpretation of the context effects on cued recall in Experiment I involves a generation–recognition model for recall (see Anderson & Bower, 1972; Bahrick, 1969; Kintsch, 1970). However, the model we have in mind, unlike the prior generation–recognition models, assumes that *S* does not recognize the words per se, but rather, semantic interpretations of these words. The *S* implicitly generates words and attempts to recognize their senses. If he recognizes the senses as being from the TBR list, then he recalls the corresponding word. The difficulty caused by a change in context is attributable to the recognition phase of recall; *S* can generate the TBR word to the strong cue but he cannot recognize it because he assigns a different interpretation to it than he assigned when the word was paired with the weak cue. In Experiment II, *S*s were instructed to generate free associations to strong cues, and then to recognize if any of the associates were TBR words. If the difficulty of recalling high-frequency words to new, strong cues results from difficulties in generation, then new, strong cues should have a greater propensity to elicit their low-frequency TBR words than would the corresponding new, strong cues to their high-frequency TBR words. However, we expected this not to be the case. Rather, we expected that high-frequency TBR words, when generated to the cues, would not be recognized as well as low-frequency TBR words.

Method

Subjects. Thirty undergraduates at the University of Michigan participated as paid *S*s. The *S*s were tested in groups of approximately 5.

Materials and apparatus. Four of the 6 lists used in Experiment I were used in Experiment II—the same 2 priming lists and 2 of the 4 critical lists. Each list consisted of 30 TBR words, again the priming lists being all medium-frequency words, while the critical lists were half high-frequency words and half low-frequency words. All words in each list were paired with a weak cue. Each list was presented by means of a Kodak Carousel slide projector.

Procedure and design. The 2 priming lists preceded the 2 critical lists. The procedure followed and the reason for including the priming lists in the design were the same as in Experiment I.

After presentation of the first of the 2 critical lists, there was the following sequence of events.

1. The *S*s were first given a sheet of paper with 30 strong extralist cues listed in a column on the left-hand side of the sheet. They were asked to generate exactly 4 free associates to each strong cue. It was pointed out to *S*s that their

free associates to a given cue might contain a TBR word from the preceding list. They were asked to circle any such words they generated.

2. After the free-association and recognition tasks, *Ss* were given a 4-alternative forced-choice (4-AFC) recognition test consisting of 30 sets of 4 words, each set containing 1 of the 30 TBR words presented in the first critical list. The 3 distractor words in each set were drawn from words generated by several individuals asked to free associate to word pairs consisting of a TBR word and its strong cue. The distractors were selected to have about the same frequency as the TBR word.

3. Finally, in order to motivate *Ss* to attend to the weak within-list cues presented in the second critical list, there was a test of cued recall involving the weak within-list cues used in the first critical list.

The presentation of the second critical list was followed by the same sequence of events, except that when *Ss* free associated, they were asked to think of as many meanings as possible for each word generated. That is, *Ss* were asked to generate 4 words to the cue, but then to try to examine each of these words for meanings not related to the cue that elicited them. This instructional manipulation was motivated by the idea that it might reduce the inhibiting effect of the semantic context induced by the strong cue on the recognition of TBR words generated by *Ss*.

Results

The data from Experiment II are presented in Table 2. The effect of instructions, i.e., to think of multiple senses of the generated words during the free-association task, was insignificant with respect to all 3 dependent variables so the data are pooled over the 2 critical lists.

As expected, the probability of generating a high-frequency word to the new, strong cue was not lower than that for a low-frequency word; if anything the reverse was true: High-frequency TBR words were generated with probability .59 to their respective cues, and low-frequency TBR words were generated with probability .55 to their strong cues. The probability that a low-frequency TBR word was recognized (.84), however, given that it was generated during the free-association task, was more than twice the probability that a high-frequency word was recognized (.38) given it was generated. The probability of recognizing low-frequency TBR words in the 4-AFC recognition test (.85) was also higher than the recognition of high-frequency words (.72), although the difference was much smaller.

TABLE 2

RECOGNITION AND CUED RECALL PROBABILITIES AS A FUNCTION OF TEST TYPE AND WORD FREQUENCY

Word frequency	Type of test		
	Generation-recognition	4-AFC recognition	Cued recall
High	.38 (.59)	.72	.61
Low	.84 (.55)	.85	.71

Note. The values in parentheses are the probabilities that a to-be-remembered word was generated in the free-association task. Abbreviation: 4-AFC = 4-alternative forced-choice recognition test.

The recognition proportions for individual *Ss* were converted via the arc sine transformation and the transformed scores were subjected to an analysis of variance with the independent variables being word frequency and method of test. The Frequency × Test Type interaction was highly significant, F (1, 29) = 30.34, $p < .001$, as was the difference between recognition of *S*-generated high- and low-frequency TBR words, F (1, 29) = 117.56, $p < .001$, and the difference between the recognition of high- and low-frequency TBR words on the 4-AFC test was also significant, F (1, 29) = 9.39, $p < .01$. Again, in consideration of possible selection artifacts, *min F'*s were computed for the Frequency × Type of Recognition Test interaction, F' (1, 73) = 16.25, $p < .001$, the difference between recognition of high- and low-frequency generated TBR words, F' (1, 87) = 41.93, $p < .001$, and the difference between frequency levels in the 4-AFC, F' (1, 50) = 7.02, $p < .025$.

The fact that the discrepancy between the recognition of high- and low-frequency TBR words is less on the 4-AFC test than it is for the recognition of *S*-generated TBR words is probably attributable to differential effectiveness of the context induced by the test. The meanings of *S*-generated TBR words are clearly determined by the strong cue from which they are generated. However, in the 4-AFC test, *Ss* seemed better able to judge

a word without regard to the other alternatives. Since the strong cue was not actually present on the 4-AFC test, it was, presumably, easier to think of the word with a meaning different from that imposed by the strong cue and more like that imposed by the original cue. Tulving and Thomson (1973) also found improved recognition of TBR words when the alternatives in the 4-AFC test were not generated by the S[4] to the strong extralist cue.

Recall to the original cues yielded results very similar to the corresponding conditions in Experiment I: The cued recall probability for high-frequency words was .61 and the corresponding probability for low-frequency words was .71.

General Discussion

The present results necessitate some refinement and clarification of Tulving and Thomson's (1973) Encoding Specificity Principle. It seems plausible, at least for retention intervals on the order of those in the present paradigm, that what is recognized in a recognition test is a particular sense of a word rather than phonemic or orthographic information. A cue may generate the same word presented earlier, but if the semantic interpretation imposed by the cue in this generation process is different from the meaning originally encoded, the word is unlikely to be recognized or recalled. It may be that if the word is tested after a brief retention interval, then phonemic cues will improve recognition. However, in experiments such as these in which Ss must use long-term memory at testing, the information retrieved would probably be a sense of the word, not its spelling pattern or phonemic characteristics per se. Past studies have shown that long-term memory confusions are semantic in nature rather than acoustic as they are in short-term memory (e.g., Craik, 1968). Furthermore, since the relation between a TBR word and its cue is a semantic one, it is even more likely in cuing experiments than in other experiments that an S would attend to the meaning of the TBR word.

Tulving and Thomson (1973) assert that the generation–recognition models for recall (Anderson & Bower, 1972; Bahrick, 1969; Kintsch, 1968, 1970) are incompatible with the findings that support their Encoding Specificity Principle. They argue that generating a word does not guarantee that it will be recognized (which is the requirement for recall) even though the "event information," as they call it, is *available*. Previous descriptions of the generation–recognition model (e.g., Anderson & Bower, 1972) stipulated that if the TBR item was encoded as occurring in the list, only implicit generation of the appropriate spelling pattern (or phonemic cues) was required for recall. Tulving and Thomson showed that this was not the case. However, if the generation–recognition models are modified by the assumption that word senses are the basis of recognition, then these models become compatible with their principle. Words, per se, are not generated to a cue, but rather senses of a word are generated. The generated meaning must match a meaning encoded during the experimental task before recall can occur.

Effect of Strong versus Weak Cues

We proposed that many words have multiple senses and frequency of the word in print is a good index of how many senses words have for a particular S. Some senses of a word will unavoidably be used and encountered more often than others. A strong cue, as manifested by its high propensity to elicit the TBR words, probably selects a salient sense of the TBR word. On the other hand, weak associations tend to be bizzarre and thus tap less likely interpretations of the TBR words. When S sees a strong cue and generates the TBR word, the sense or interpretation which comes to mind is totally different from the obscure one studied, and S frequently cannot think of all the possible meanings related to a word generated and thus does not "notice" the studied sense. Therefore, it would seem that the more obscure or removed the old interpretation, the less likely that it will be noticed, and consequently, the less likely the word would be recognized.

Tulving and Thomson noted that their theory could not account for the finding that weak input cues paired with strong output

[4] The difference between recognition of words generated by S and recognition of words that were generated by a yoked S was in the same direction as our results although the difference was smaller. The smaller difference may be due to the fact that in their experiment the strong cue words were presented along with the distractors.

cues produced higher recall than strong input cues and weak output cues, as they expected no difference (Thomson & Tulving, 1970). If one assumes, however, that recall requires both generation and recognition, then one should expect better recall when the output cues are strongly associated rather than weakly associated because, in the task of recall, the candidates must be thought of before they can be recognized as words from the list. There is simply a much higher probability of *generating* the TBR word with the strong cue, regardless of the 2 cues' relative effects on recognition. Therefore, one would have to predict that weak input cues with strong recall cues give better performance than the converse.

Given the generation–recognition distinction and the assumption that it is senses that are recognized, the predictions for recognition are the opposite of those made for recall: A TBR word presented with a strong cue and later tested with a weak cue should be better recognized than a word originally presented with a weak cue and later tested with a strong cue. This prediction follows because the TBR word in the presence of a strong associate is not likely to suggest the obscure alternate sense of the word studied during its presentation with a weak cue. On the other hand, if the TBR word were originally presented with a strong cue, a frequent interpretation of the word would be studied. Because this sense is so salient, *S* would be likely to revive that same sense even when the word is presented in the context of a new, weak associate. This prediction was also confirmed in Tulving and Thomson's (1971) data, but no explanation was offered for that outcome.

Thomson (1972) also did not explain a result from one of his recognition experiments in which context was deleted. He found that presenting only 1 of 2 words originally presented as a pair had a less deleterious effect when the original pair had a strong associative relation than when it had a weak one. Also, the most harmful effect on recognition performance occurred on tests of single words initially presented on the right side of weakly associated input pairs. One possible explanation is that strongly related pairs tend to be encoded with the common sense of each word, so if either of the words is presented separately, the original sense is very likely to be noticed. On the other hand, when 2 words are only weakly related, at least 1 of the 2 words must be encoded in terms of a less frequently used sense of the word. It seems reasonable that the first word encountered when reading the pair imposes its dominant sense on the second word (unless, for some reason, a more obvious interpretation employs a common sense of the second and an obscure sense of the first). Thus, it seems plausible that a word on the right in a weakly associated pair would be likely to have an infrequent and unusual meaning encoded, which, in turn, would produce poor subsequent recognition of the word.

Effects of Word Frequency

As stated earlier, we think that word frequency is a better index of the average number of senses of a given word than is number of dictionary meanings. Although the 2 indices are strongly correlated for the words in this experiment ($r = .61$), the frequency measure correlated better with the obtained results. Using the data of Experiment I, a difference score was computed for each TBR word by taking the number of correct recalls to the original weak cue and subtracting the number of correct recalls to the new, strong cue. This difference score correlated .37 with frequency and .17 with number of dictionary meanings. The correlation between number of dictionary meanings and the difference score when frequency was partialed out was −.06 and frequency correlated .33 with the difference score when the number of dictionary meanings was partialed out. The first partial correlation is negative and not significant, $t\ (117) = .65$, $p < .60$. The second is significant, $t\ (117) = 3.78$, $p < .001$.

Rubenstein, Garfield, and Millikan (1970) provide incidental support for the notion that high-frequency words are basically quite similar to homographs and that frequency may be a better indicator of number of senses in a person's lexicon than dictionary entries. They administered a task that showed that decisions concerning whether letter strings are words had shorter latencies for homographs than nonhomographs. Their explanation was that homographs have more lexical entries and hence the probability of matching the stimulus with an entry is greater in any period of time, thus making reaction times shorter for homographs. Rubenstein et al. also found a highly significant effect for frequency of the word, high-frequency words being faster than lower frequency words; in fact, in a reanalysis of their data, Clark (1973)

found that the homography effect per se was insignificant, but that the strong frequency effect remained. Neither Rubenstein et al. nor Clark considered the notion that frequency might be a very sensitive measure of the number of meanings for a given word in a particular *S*.

REFERENCES

ANDERSON, J. R., & BOWER, G. H. Recognition and retrieval processes in free recall. *Psychological Review*, 1972, 79, 97–123.

BAHRICK, H. P. Measurement of memory by prompted recall. *Journal of Experimental Psychology*, 1969, 79, 213–219.

BOBROW, S. A. Memory for words in sentences. *Journal of Verbal Learning and Verbal Behavior*, 1970, 9, 363–372.

CLARK, H. H. The language-as-fixed-effect fallacy: A critique of language sampling in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 1973, 12, 335–359.

CRAIK, F. I. M. Types of error in free recall. *Psychonomic Science*, 1968, 10, 353–354.

KINTSCH, W. Recognition and free recall of organized lists. *Journal of Experimental Psychology*, 1968, 78, 481–487.

KINTSCH, W. Models for free recall and recognition. In D. A. Norman (Ed.), *Models of human memory*. New York: Academic Press, 1970.

KUČERA, H., & FRANCIS, W. N. *Computational analysis of present-day American English*. Providence: Brown University Press, 1967.

LIGHT, L. L., & CARTER-SOBELL, L. Effects of changed semantic context on recognition memory. *Journal of Verbal Learning and Verbal Behavior*, 1970, 9, 1–11.

MACKAY, D. G., & BEVER, T. A. In search of ambiguity. *Perception & Psychophysics*, 1967, 2, 193–200.

PAIVIO, A., YUILLE, J. C., & MADIGAN, S. A. Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monograph Supplement*, 1968, 76(1, Pt. 2).

RUBENSTEIN, H., GARFIELD, L., & MILLIKAN, J. A. Homographic entries in the internal lexicon. *Journal of Verbal Learning and Verbal Behavior*, 1970, 9, 487–494.

SCHNORR, J. A., & ATKINSON, R. C. Study position and item differences in the short- and long-term retention of paired associates learned by imagery. *Journal of Verbal Learning and Verbal Behavior*, 1970, 9, 614–622.

THOMSON, D. M. Context effects 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 wordbook of 30,000 words*. New York: Columbia University, Teachers College, Bureau of Publications, 1944.

TULVING, E. Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory*. New York: Academic Press, 1972.

TULVING, E., & THOMSON, D. M. Retrieval processes in recognition memory: Effects of associative context. *Journal of Experimental Psychology*, 1971, 87, 116–124.

TULVING, E., & THOMSON, D. M. Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 1973, 80, 352–373.

Webster's seventh new collegiate dictionary. (7th ed.) Springfield, Mass.: G. & C. Merriam, 1965.

WINOGRAD, E., & CONN, C. P. Evidence from recognition memory for specific encoding of unmodified homographs. *Journal of Verbal Learning and Verbal Behavior*, 1971, 10, 702–706.

(Received July 14, 1973)