

Encoding specificity and recognition memory for words

STEVEN SCHWARTZ

University of Texas Medical Branch, Galveston, Texas 77550

The question of whether performance, in recognition memory tasks, depends upon the recognition of a specific interpretation of a word (as originally encoded) or upon its physical (orthographic-phonemic) representation was investigated using a model of the recognition process based on signal-detection theory. Thirty-six subjects studied lists of common and rare words. Each word was paired with a weakly associated cue. On a subsequent recognition task, old words were presented along with strongly associated cues. This change in accompanying cue impaired recognition for common but not for rare words. Differential recognition rates were not, however, associated with different levels of retention (as measured by d'). Rather, the effect was due to the cautious decision criterion employed for responses to common word meanings. The data suggest that recognition of a word depends upon its physical representation as well as its meaning and that the latter functions so as to affect response biases and not retention.

It is quite clear that the ability to recognize a previously studied word is highly dependent on both the context in which the word was originally presented and the context present during testing (e.g., Light & Carter-Sobell, 1970; Tulving & Thomson, 1971). In an attempt to explain the effects of context, Tulving and Thomson (1973) introduced an "encoding specificity principle," which asserts that the memory trace of an input event (and, therefore, the characteristics of a good retrieval cue) is determined by the specific encoding procedures performed on the input stimuli. Within this scheme, to-be-remembered (TBR) items are thought to be stored as higher order episodic units and only elements of the episodic unit are thought capable of accessing the unit (see Tulving, 1972).

A recent paper by Reder, Anderson, and Bjork (1974) reported the results of experiments designed to clarify some of the possible interpretations of the encoding specificity principle. In both recall and recognition studies, they found memory to be adversely affected by changes in the cues accompanying common TBR words. This effect was obtained when the change was from a cue suggesting an obscure meaning of the word to one suggesting a common meaning. Memory for rare words, on the other hand, was unaffected by changes in the accompanying cues. Since rare words were thought to possess relatively few meanings (which are, therefore, unchanged by changes in their accompanying cues), this result was taken as support for the hypothesis that memory involves the retention of a particular "sense" of a word rather than its phonemic or orthographic characteristics. There exists, however, convincing evidence that the sensory features of verbal material do persist for some time after presentation and that these features affect performance, at least on recognition tasks (e.g., Kirsner, 1974). One is forced to conclude,

then, that both types of information (sensory and semantic) may, in certain situations, influence recognition memory for words. This does not mean that both types of information are equally well retained. That is, accuracy (particularly in recognition tasks) depends not only on the type and amount of information retained but also on the contribution of response biases to the decision-making process. In the experiments reported by Reder et al., for example, all words were paired with weak cues during learning in order to give common words obscure interpretations. As rare words, or at least rare meanings, were all that were encountered during learning, subjects may have been reluctant to give common word meanings as responses during testing. Such a response bias could have resulted in greater recognition accuracy for rare words, not because such words were better retained (because their meanings were unchanged), but because of a cautious decision criterion for common words at the time of testing. This interpretation of Reder et al.'s finding is supported by the results of an experiment reported by Hunt and Ellis (1974) in which only common word meanings were employed. They found that changing the cues accompanying TBR words during testing from those present during learning resulted in the same amount of recognition impairment regardless of whether the TBR word's meaning was changed or not.

In order to determine the relative effects of semantic and sensory-temporal features on retention and response biases respectively, it is necessary to employ a basic model which allows the separation of the two. Several such models exist. The one adopted here is derived from signal-detection theory (Tanner & Swets, 1954) and is applied to the recognition memory task. It is assumed that test items possess a "familiarity" value which varies in a Gaussian fashion and which is greater for TBR

words than for new words. The differences between the familiarity of presented and nonpresented items (d') provides a measure of retention independent of response bias. Response bias is conceived of as the position of the criterion applied to the familiarity axis in order to determine the response. The research support for this approach to recognition memory is discussed and summarized by Broadbent (1971).

The present research employed signal detection parameters in order to clarify the precise nature of the information retained in memory recognition tasks. It was hypothesized that if it is actually a semantic sense of the word that is retained, we should observe a lower sensitivity (d') for words whose meanings are changed relative to words whose meanings have not changed with a change in an accompanying cue.

METHOD

Subjects

Thirty-six undergraduate psychology students served as subjects. They received course credit for participating and were run in small groups of four.

Materials

Two practice and four experimental lists were used in this study. The practice lists contained 28 medium-frequency words, each paired with a common associate taken from the Bilodeau and Howell (1965) list. The experimental lists each contained 28 words (14 high- and 14 low-frequency words) selected so that half of the words were rated A or higher on the Thorndike-Lorge (1944) norms and half of the words had frequencies of less than 10/1,000,000. All of the TBR words were high meaningful nouns (Paivio, Yuille, & Madigan, 1968). Each word was paired with a weak cue. Weak cues were words judged to elicit the TBR words infrequently. Due to the difficulties involved in establishing norms for low-frequency responses to stimulus words, cues were generated intuitively according to the criteria set forth by Reder et al. (1974). That is, care was taken to insure that strong cues for low-frequency words were not more effective than strong cues for high-frequency words. A pretest showed high-frequency words to be given as associates to their cues more often than low-frequency words were given as associates to their cues and that no two cues to a given word were synonyms. The word pairs appeared side by side with the TBR words to the right in upper-case and the cues to the left in lower-case letters.

Procedure and Design

The lists were presented to the subjects by a slide projector for 2.5-sec exposures. Subjects were told that they must try to remember the capitalized word and that the cue word could help them to do so. The practice lists were included in order to get the subjects used to using the cue words. The two practice lists preceded the experimental lists. After each list, the subjects were shown a series of word-pair probes and asked to judge whether the word in capital letters appeared earlier or whether it was a new word. Subjects were asked to record their confidence in their judgments on a scale that ranged from one (certain this word is old) to five (certain this word is new). For each experimental list, subjects were tested with 28 probes: 14 old words (seven rare and seven common) chosen at random from the original list, each accompanied by a new, strongly associated cue; and 14 new words (seven rare and seven common) chosen from the frequency classes already described and each accompanied by a strongly associated cue. The presentation

order of both the experimental lists and of the probes for each list was varied randomly across testing sessions. Ten seconds was allowed between probes for subjects to record their confidence judgments in answer booklets provided for this purpose.

RESULTS

The percentages of correct recognitions (hits) of common and rare words were .69 and .80, respectively. The difference between these two hit rates was statistically significant, $t(35) = 2.05$, $p < .03$. The false alarm rates of common and rare words were .30 and .34, respectively. The difference between these two false alarm rates was not statistically significant ($p > .05$).

In order to determine whether this difference in correct recognitions reflected differences in sensitivity, d' was estimated for both rare and common words. Due to the relatively small number of judgments each subject was asked to make, there was little evidence from false positives for some subjects in the "certain" category. For this reason, each session's data were pooled across subjects. Since subjects participated in groups of four, the experiment required nine sessions. For statistical analyses, then, the unit was sessions, and the sample was nine. The difference between the mean values of d' taken at the most risky criterion (1.04 and 1.21 for common and rare words, respectively) approached but did not reach statistical significance ($p > .05$). Since d' did not differ, it seemed plausible that differences in criterion (β) may have been responsible for the superior recognition of rare words. Following the procedures outlined by Broadbent (1971, pp. 70-76), β for both types of words was also calculated at the most risky category. The difference between these mean values of β (1.03 for common and .76 for rare words) was statistically significant, $t(8) = 2.03$, $p < .05$.

DISCUSSION

The results of the present study are in accord with those reported by Reder et al. (1974). Changes in the cues accompanying TBR words (from weak to strong cues) adversely affects recognition memory for common, but not for rare, words. A signal-detection analysis, however, did not provide evidence that this effect was due to differential retention. Rather, the effect seemed to be due to a selective response bias. Subjects exposed to only rare words (or at least rare meanings) during learning seemed reluctant to use common word meanings as responses even though the words were remembered as well as rare words.

The findings of this study have some important implications for some of the results of Reder et al. not yet discussed. In their second experiment, two types of recognition tasks were conducted. The first required that the subject produce associates to strong, extra-list cues and then to indicate which, if any, of these generated words appeared on the original list. The second test, a four-alternative, forced-choice test, required that subjects choose old words from among one TBR word and three associates. Although recognition for common words was poorer than for rare words on both tests, the difference between the two types of words was much smaller on the forced-choice test than on the generation-recognition test. This sort of task

difference is quite predictable in the light of the present findings. The generation-recognition test did not require that the subject make a response on each trial and, as a consequence, a conservative subject could opt not to respond on those trials containing common words (when he held a cautious criterion). Such a subject would be likely to miss many common words. The forced-choice task, on the other hand, required that subjects choose a response on each trial (in effect, subjects were forced to behave in a risky fashion). As expected, such a procedure improved accuracy for common words by a factor of almost two.

The results of the present study are in substantial agreement with a growing body of evidence that, in recognition, at least, physical information may persist for some time after original learning. Moreover, these results suggest that under certain circumstances, phonemic and orthographic information may be as useful in a recognition task as the particular meaning of a word (see also, Nelson & Brooks, 1974).

REFERENCES

BILODEAU, E. A., & HOWELL, D. C. *Free association forms*. (Catalog No. D 210.2:F 87). Washington, D. C: U. S. Government Printing Office, 1965.
 BROADBENT, D. E. *Decision and stress*. New York: Academic Press, 1971.
 HUNT, R. R., & ELLIS, H. C. Recognition memory and degree of semantic contextual change. *Journal of Experimental Psychology*, 1974, **103**, 1153-1159.

KIRSNER, K. Modality differences in recognition memory for words and their attributes. *Journal of Experimental Psychology*, 1974, **102**, 579-584.
 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.
 NELSON, D. L., & BROOKS, D. H. Relative effectiveness of rhymes and synonyms as retrieval cues. *Journal of Experimental Psychology*, 1974, **102**, 503-507.
 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).
 REDER, L. M., ANDERSON, J. R., & BJORK, R. A. A semantic interpretation of encoding specificity. *Journal of Experimental Psychology*, 1974, **102**, 648-656.
 TANNER, W. P., JR., & SWETS, J. A. A decision-making theory of visual detection. *Psychological Review*, 1954, **61**, 401-409.
 THORNDIKE, E. L., & LORGE, I. *The teacher's wordbook of 30,000 words*. New York: Columbia University Press, 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. *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.

(Received for publication June 4, 1975.)