

# Predicting recognition during storage: The capacity of the memory system to evaluate itself

LOWELL D. GRONINGER

*University of Maryland Baltimore County, Baltimore, Maryland 21228*

The major goal of this experiment was to determine the accuracy with which subjects could predict at input their later recognition performance. Predictions were based on words varying on a number of attributes. Four categories of words were used: nonsense, abstract, concrete, and emotional. The number of words correctly recognized increased in the foregoing order although the difference between the concrete and emotional words was marginal. When all words were combined, there was a strong relationship between predicted recognition and actual recognition for words that were hits with a "very sure" confidence rating. Additionally, subjects considerably underestimated their later recognition performance. The results indicated that a new interpretation of the mechanism underlying "correct rejects" is feasible. The relevance of the results to other memory issues was also discussed.

Despite the voluminous literature on verbal learning and memory, much remains to be learned about what would seem to be a core issue: What happens in terms of cognitive variables when a person sees or hears a word? A greater appreciation of this deceptively simple statement is obtained if a few possibilities are outlined. Words are not stored solely on the basis of their phonological characteristics, rather a storage complex in the sense of several attributes (Bower, 1967; Morton, 1970; Norman & Rumelhart, 1970; Underwood, 1969, 1972; Wickens, 1970, 1972) or propositions (Anderson & Bower, 1973) seems to be a more appropriate description. But is this complex automatically activated given a specific context or does conscious control play a role (Posner & Warren, 1974)? Consider another problem. How can a person accurately predict that he knows a word, in the absence of recall of that word—the "tip of the tongue" phenomena (Brown & McNeill, 1966)?

The ramification of these questions about memory structure in relation to recognition memory was the focus of the present investigation. More specifically, consider the remarkable ability of subjects to correctly reject words not previously presented. When subjects place a high degree of confidence in their decisions, they are correct about 80%-90% of the time, the exact percentage depending on the usual variables affecting recognition memory. The traditional method of accounting for correct rejects as well as hits is to use a familiarity dimension, the subject reporting that he is sure that a word is "old" if it is very familiar (within the context

of the experiment) and sure that a word is "new" if it is unfamiliar. One explanation of the mechanism underlying familiarity judgments comes from what Tulving and Thompson (1973) call generation-recognition models of memory. Within this framework, attributes of a word determine the word's location in memory. A judgment of "old" or "new" then depends on whether or not the location determined by the attributes of a given word has been recently activated or "tagged."

This analysis of the memory system assumes that the system is passive in the sense that representations of words are automatically activated and evaluations of the words are made after this activation. Although this may well happen for some words, the suggestion here is that the basis for judging many words as "new" is neither from an activated empty location nor from the lack of an appropriate context at retrieval, as the theory of encoding specificity (Tulving & Thompson, 1973) would predict. Rather, judgments are based on an active analysis of the memorial quality of the words during input. If a word is judged as easy to remember, then a subject knows that a "new" word is "new," because if it were "old" he knows that he would have remembered it. For words judged as not easy to remember, other factors such as context become more important. However, before this analysis can be considered tenable it must first be shown that subjects can accurately predict what they will recognize.

To minimize problems concerning end points on a subjective scale involving predictions of the later recognition of words, it was decided to use four categories of words ranging from nonsense "words" to emotional words. From these categories arose a secondary goal of the present investigation. This goal was to test the hypothesis that the recognition of a word depends, in part, on the number of different attributes, broadly defined,

Requests for reprints should be sent to Lowell D. Groninger, Department of Psychology, University of Maryland Baltimore County, Baltimore, Maryland 21228. Marilyn D. Wang sponsors this paper and takes full editorial responsibility for its contents.

associated with that word. Thus the components of words were increased from no direct meaning, to meaning, to meaning plus images, to meaning plus images plus affect. On the basis of studies by Elias and Perfetti (1973), Gorman (1960), Groninger (1974), and Sadulka and Loftus (1972), it was anticipated that recognition would increase in the order of the conditions mentioned. At a minimum, it was felt that this dimension would present end points for the prediction ratings.

## METHOD

### Experimental Design

The design was a within-subjects design wherein a matrix was composed comparing subjects' initial ratings on their confidence that they would recognize a word with their later recognition performance. Four different types of words were presented within these conditions. These word types were: *nonsense*—no dictionary meaning and no obvious (rated by three judges) confusability in sound to a word; *abstract*—a rating of below 4.5 on the Paivio, Yuille, and Madigan (1968) norms; *concrete*—a rating of above 6.0 on the Paivio et al. (1968) norms; and *emotional*—words chosen by the same three judges to have sexual connotations. No slang words were used in the lists.

### Procedure

Fifty words from each category were selected by the criteria previously stated. Thirty words from each category were then arranged in random fashion throughout the first list, with the constraint that 40 words, 10 from each category, would appear in the first 30 and last 10 positions in the list. These first and last words were not included in the data analysis, in order to largely eliminate a serial position effect. Thus, the ratings reflect the most difficult part of the list. An additional reason for eliminating these words was to let the subjects evaluate several words to get a "feel" for the task. Giving what amounts to 30 practice words should stabilize the ratings of the remainder of the words. For the recognition memory test the 80 remaining target words were randomly interspersed with 80 distractors, 20 from each of the four categories.

The words were presented at a 3-sec rate and tested at a 5-sec rate. Subjects were told at the initial presentation to rate each word on a 5-point bipolar scale ranging from "very sure will recognize" to "very sure won't recognize." They were also told at the time that they would later be given a recognition memory task wherein they would be required to distinguish these words from others. The scale used for the recognition memory test was a bipolar 5-point scale ranging from "very sure the word was on the first tape" to "very sure the word was not on the first tape." Subjects participated in the experiment in groups of 10 or fewer. Subjects were seated about 3 ft apart and listened to the words, played from a Sony stereo tape-recorder, in a large room. After the presentation of both tapes, subjects were asked to rate on a bipolar 5-point scale their attentiveness to the words throughout the experiment. The experimental situation, including the presence of an experimenter, made copying ratings among subjects a negligible factor.

### Subjects

The subjects were 69 members of the introductory psychology class at the University of Maryland Baltimore County who were fulfilling part of a laboratory requirement. Subjects were chosen without regard to sex.

**Table 1**  
Hits and False Alarms for Data With  
the Highest Confidence Rating

		Word Type			
		Emotional	Concrete	Abstract	Nonsense
Hits	Mean	15.27	14.33	11.30	4.73
	SD	4.24	4.10	4.21	3.60
False Alarms	Mean	1.22	.50	1.45	.85
	SD	1.77	.88	1.80	1.43

Note. Scores are based on 20 target words and 20 distractor words per category.

## RESULTS

The data from the two subjects who reported that their attentional levels were poor or very poor were discarded from further analysis. The remaining N on which the analysis was made was 67. The data were then sorted into confidence ratings for predicted recognition and scored for hits and false alarms by word type. Since the pattern of results was the same with or without the data with the "fairly sure" ratings, the data in Table 1 reflect only the highest confidence ratings, the "very sure" rating. One advantage of basing the analysis only on the "very sure" data is that the false alarms are thereby minimized. Minimizing false alarms is desirable because there is no direct way to adjust for false alarms in the predicted recognition vs. actual recognition matrix shown in Table 2.

The results in Table 1 are consistent with what would be predicted if words are identified from the number of attributes associated with them. The nonsense "words," having neither direct meaning nor direct referents, had the lowest recognition scores. They were probably identified by either their sounds or some mediating link to a word. The abstract-concrete difference was significant at the .01 level using a t test. This difference has been established in other studies such as Gorman (1960). The high scores of the emotional (sexual) group of words was also anticipated; however, the confounding resulting from these words forming a category of limited size prevents definitive uncomplicated conclusions. The emotional words have a slightly higher, though nonsignificant,  $p > .05$  hit rate compared to the concrete words. Their false alarms were also higher,  $t(112) = 4.29$ ,  $p < .01$ . The higher false alarm rate of the emotional words is not surprising since the distractors in this category were much more similar to the targets than in the concrete category.

The primary intent of the word classifications was to establish a broad scale by which words could be contrasted, thus maximizing whatever relationship might exist between predicted recognition and actual recognition. The data showed this expected contrast, with the nonsense "words" clustering at the low end of the predictive scale, the emotional words clustering at the high end, and the others clustering in the middle. As might

Table 2  
 Predicted Recognition vs. Actual Recognition by Confidence Rating

Predicted Recognition		Recognition Confidence Rating				
		Hit	Midpoint			Miss
		Very Sure	Fairly Sure	Don't Know	Fairly Sure	Very Sure
Very Sure Will	Mean	18.88	1.50	1.08	.95	.92
	Mean/Row	.81	.05	.05	.04	.14
	Mean/Column	.40	.11	.11	.14	.10
Fairly Sure Will	Mean	12.40	2.26	1.44	1.44	.65
	Mean/Row	.68	.12	.08	.08	.04
	Mean/Column	.27	.19	.15	.21	.13
Don't Know	Mean	7.83	2.80	1.94	1.56	1.05
	Mean/Row	.52	.18	.13	.10	.07
	Mean/Column	.17	.24	.20	.23	.22
Fairly Sure Won't	Mean	4.55	2.64	2.55	1.70	1.10
	Mean/Row	.37	.21	.20	.13	.09
	Mean/Column	.10	.23	.26	.25	.23
Very Sure Won't	Mean	2.94	2.44	2.73	1.23	1.17
	Mean/Row	.28	.23	.26	.12	.11
	Mean/Column	.06	.21	.28	.18	.24

Note. Data are based on 80 word list.

also be expected, the relationship between predicted recognition and actual recognition within the various classification of words was in the same direction but not as large as the same relationship resulting from the combination of all words.

All categories of words were combined to form the data in Table 2. These data are presented in terms of cell means and the proportion of cell means to the row and column totals. The proportion involving rows indicates predictive accuracy, while the proportion involving columns indicates the extent to which a particular prediction exhausts a particular confidence rating set at retrieval. For example, the first cell shows that when subjects are very sure they will recognize a word, the probability is .81 that they will recognize that word with a confidence rating of "very sure." The first cell also shows that, of the words recognized with a "very sure" rating, 40% were words given a prediction rating of "very sure will recognize."

Since differences from predictions showed up in nontrivial form only in hits with "very sure" ratings, only these data were subjected to statistical tests. The paired comparison of means in this column in descending order were all statistically significant by *t* tests at the .01 level. However, the really interesting and important finding is in the accuracy of predictions by confidence rating. When subjects are very sure that they will recognize a word, they are about six times as likely to correctly recognize that word with a confidence rating of "very sure," as contrasted to their being very sure that they will not recognize that word. That the subjects underestimated their abilities can best be seen by observing the high proportions of hits even when the predictions of recognition were marginal or negative.

## DISCUSSION

There are several important aspects of the data in relation to storage phenomena. In terms of prediction, the data in Table 2 seem nearly dichotomous, with the predictability of recognition showing up almost exclusively in words that were later hits with a "very sure" confidence rating. The predictability for all other confidence ratings at recognition is only slight in comparison. Thus, only words that get processed into the system very distinctly, as shown by the later recognition memory test, show the relationship between prediction and later recognition. Additionally, the predictions when subjects were not sure or thought that they would not recognize the words were considerably underestimated. It is as though two types of word attributes are operating, only one of which subjects can utilize in making correct predictions. One class of attributes allows recognition of words without subjects being aware at initial presentation that these attributes will enable later recognition. The other class of attributes is identified at input as salient for recognition, and this prediction is later confirmed. The gist of this interpretation is consistent with accounts of attribute formation suggested by Posner and Warren (1972) and Wickens (1972). Both suggest some attribute encoding prior to conscious processing. While the present data cannot confirm that some attributes do not reach consciousness, it does suggest that some attributes, if they do reach consciousness, are not evaluated according to their full potential for subsequent retrieval.

The present data are also relevant to what is termed "the feeling of knowing" phenomenon (Hart, 1967; Blake, 1973). This is a retrieval phenomenon involving the prediction of the availability of words in memory in the absence of the recall of those words. The present experiment involves the prediction of the later recognition of words, at the time of their presentation—a seemingly more difficult task since the prediction is made during storage rather than during a retrieval attempt. Blake (1973) interprets this "feeling of knowing" phenomenon in terms of the identification of some of the attributes of an item, but not the attributes which directly lead to the verbal label. However, enough information is available to be confident that the item can be identified given sufficient cues. The present data support the assumptions involved in this interpretation by demonstrating that the prediction of recognition can be very good and that

actual recognition depends, in part, upon the variety of attributes associated with that item.

As previously suggested, a further byproduct of the accuracy of predicting hits, as found in the present study, is in suggesting a new interpretation of the familiarity dimension, often referred to in recognition memory studies as the basis upon which "old" and "new" decisions are made. While "old" judgments may be based on a familiarity basis of some type, a confident "new" judgment seems more likely to be based on the memorable qualities of the attributes of words. An example that takes this argument to its limit is that a student may not know if the straight-laced psychology professor mentioned William James in his lecture, but he knows that he did not use any vulgar slang terms, because he knows that he would have remembered it had the professor done so. One ramification of this position is that the potential bias from distractors in recognition memory studies is very large. A bias can enter not only from an overlap in attributes between target words and distractors, but also from very high predictive indices of distractors which prevent them from becoming reasonable candidates for the target list.

Finally, Underwood (1969) and Tulving and Madigan (1970) imply that determining the capacity of the memory system to appraise its own storage is of fundamental significance. This capacity has been demonstrated in the present study in its most difficult form, an appraisal of storage during storage.

#### REFERENCES

- ANDERSON, J. R., & BOWER, G. H. *Human associative memory*. New York: Wiley, 1973.
- BLAKE, M. Prediction of recognition when recall fails: Exploring the feeling of knowing phenomenon. *Journal of Verbal Learning and Verbal Behavior*, 1973, 12, 311-319.
- BOWER, G. H. A multicomponent theory of the memory trace. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 1). New York: Academic Press, 1967. Pp. 229-325.
- BROWN, R., & MCNEILL, D. The "tip-of-the-tongue" phenomenon. *Journal of Verbal Learning and Verbal Behavior*, 1966, 5, 325-337.
- ELIAS, C. S., & PERFETTI, C. A. Encoding task and recognition memory: The importance of semantic encoding. *Journal of Experimental Psychology*, 1973, 99, 151-156.
- GORMAN, A. M. Recognition memory for nouns as a function of abstractness and frequency. *Journal of Experimental Psychology*, 1961, 61, 23-29.
- GRONINGER, L. D. *The role of images in recognition memory*. Unpublished manuscript, University of Maryland, Baltimore County, 1974.
- MORTON, J. A functional model for memory. In D. A. Norman (Ed.), *Models of human memory*. New York: Academic Press, 1970.
- NORMAN, D. A., & RUMELHART, D. E. A system for perception and memory. In D. A. Norman (Ed.), *Models of human memory*. New York: Academic Press, 1970.
- PAIVIO, A., YUILLE, J. C., & MADIGAN, S. A. Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monograph*, 1968, 76(Pt. 2).
- POSNER, M. I., & WARREN, R. E. Traces, concepts and conscious constructions. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory*. New York: Wiley, 1972.
- SADULLA, E. K., & LOFTNESS, S. Emotional images as mediators in one-trial paired-associate learning. *Journal of Experimental Psychology*, 1972, 95, 295-298.
- TULVING, E., & MADIGAN, S. A. Memory and verbal learning. *Annual Review of Psychology*, 1970, 21, 437-484.
- TULVING, E., & THOMPSON, D. M. Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 1973, 5, 352-373.
- UNDERWOOD, B. J. Attributes of memory. *Psychological Review*, 1969, 76, 559-573.
- UNDERWOOD, B. J. Are we overloading memory? In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory*, New York: Wiley, 1972.
- WICKENS, D. D. Encoding categories of words: An empirical approach to meaning. *Psychological Review*, 1970, 77, 1-15.
- WICKENS, D. D. Characteristics of word encoding. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory*. New York: Wiley, 1972.

(Received for publication January 19, 1976.)