Recognition memory for words, pictures, and words + pictures

JOAN GAY SNODGRASS, ROCHELLE VOLVOVITZ
and ELEANOR RADIN WALFISH
New York University, University Heights, Bronx, N.Y. 10453

Using the method of signal detection theory (SDT), recognition memory was measured in two groups of Ss, with one group seeing words alone and pictures alone and the other group seeing words alone and pictures + words. Both groups were run for three sessions under each of three presentation probability conditions. Recognition memory was better for pictorial stimuli in each group, but there was no improvement in recognition memory for the combined cues of word + picture over picture alone. Four measures of recognition memory, including two based on SDT and one based on high threshold theory, were highly correlated.

The present study is concerned with memory for two different kinds of stimuli—words and pictures, and their combination—in a recognition memory paradigm. Two questions are posed: Is memory for pictorial material better than for verbal material? How does performance for their combined cues compare with performance for each separate cue?

There is a good deal of evidence that, for a variety of tasks, pictorial imagery variables play an important role. For example, words rated higher in imagery are remembered better than those rated low in imagery (Paivio, 1969), and recognition memory for pictures is better than that for words (Shepard, 1967).

One would also expect that presenting two cues in a recognition memory paradigm, such as the combination of a word with its pictorial illustration, would improve recognition memory even more. For example, if the cues were stored independently, one would expect predictably better performance with two cues than with one. Furthermore, Tulving & Thomson (1971) have shown that the presentation of two cues present during learning improves recognition memory over the presentation of a single cue, supporting the view that recognition involves retrieval operations.

SUBJECTS

Sixteen Ss (eight males and eight females), all undergraduates of New York University, served in the experiment. Eight Ss served in Group A, in which recognition memory for words was compared with recognition memory for pictures, and eight Ss served in Group B, in which recognition memory for words was compared with recognition memory for words + pictures. Half of the Ss in each group were female.

APPARATUS AND PROCEDURE

The basic procedure was identical for both groups of Ss. Using a procedure first introduced by Egan (1958), a recognition memory task was carried out using the methodology of signal detection theory (SDT). Ss were presented with an inspection pack of 60 cards, exactly half of which were words alone and half either pictures alone (for Group A Ss) or pictures + words (for Group B Ss). The pictures were drawings illustrating the words, and they appeared with their appropriate word in the word + picture condition. Ss in both groups received the same words alone and the same pictures. The only difference between stimuli experienced by Group A and those experienced by Group B was that the drawings were accompanied by the words they illustrated for Group B and appeared alone for Group A. Of the words in the word alone set, 91% were rated AA or A in the Thorndike & Lorge (1944) counts, and 9% had counts ranging from 48 to 8 per million words; words chosen to be illustrated for the picture alone or the word + picture condition consisted of 86% AA or A words and 14% with frequencies between 48 to 8 per million words. Ss were permitted to view the 60 items for 2 min. After the inspection period, S was given a 2-min rest period, which was followed by the recognition test period, during which S was presented with a shuffled pack containing both old and new items. The new items consisted of words alone and pictures alone for Group A and of words alone and words + pictures for Group B. S's task was to separate the items into those which were old and those which were new. He was permitted as much time as he wished during this part of the experiment.

Each S served in three sessions, during which he experienced a different set of words, pictures, or words + pictures. Each S received each of the three signal presentation probabilities, 0.4, 0.5, and 0.6. The order of presentation probabilities and the particular set associated with each probability were determined randomly for each S. The presentation probability was varied by varying the number of new items introduced in the test session with the 60 old items, Thus, 90, 60, or 40 new items were added during the recognition test for the 0.4, 0.5, and 0.6 conditions, respectively. Each S was informed at the beginning of the recognition test period of the presentation probability condition in effect, and Ss were motivated to earn a maximum number of points by the offer of a prize of $3.00 to the S with the largest number of points. A symmetric payoff matrix awarding 1 point for hits and correct rejections and imposing a penalty of 1 point for false alarms and misses was employed throughout the experiment. The purpose of the presentation probability manipulation was to manipulate S's criterion so that an ROC curve could be fit to the data points, and from it d' (the SDT measure of sensitivity, or, in this context, memory strength) calculated. This is similar to the approach of Murdock (1965), except that he used rating scales rather than variation in presentation probabilities.

RESULTS

The overall hit and false alarm rates across Ss for both Group A and Group B Ss are shown in Table 1. It is apparent that presentation probability had a small and inconsistent effect on response bias; thus, it was not possible to calculate SDT measures of d' from ROC curves. Instead, d' was calculated from the combined hit and false alarm rates for each presentation probability condition. In addition, three other measures of recognition memory were calculated. These were the probability of a correct response, P(C); the measure of true probability of hits, P*(H), derived from high threshold theory, given by:

$$P^*(H) = \frac{P(H) - P(FA)}{1 - P(FA)}$$

Table 1
Hit and False Alarm Rates for Groups A and B for Each Presentation Probability Condition

Presentation Probability		Group A		
		0.4	0.5	0.6
Words Alone	H	.688	.721	.771
	FA	.200	.225	.300
Pictures Alone	H	.850	.812	.817
	FA	.069	.054	.050
		Group B		
Words Alone	H	.692	.775	.708
	FA	.117	.188	.206
Words + Pictures	H	.800	.725	.833
	FA	.042	.058	.031

Table 2

Table 2
Four Measures of Recognition Memory: d', Probability Correct, P*(H), and A' for Each Group and Each Presentation Probability*

Presentation Probability		Group A		
		0.4	0.5	0.6
Words Alone	d'	1.33	1.34	1.27
	P(C)	.744	.748	.736
	P*(H)	.610	.640	.673
	A'	.830	.832	.821
Pictures Alone	d'	2.52	2.49	2.55
	P(C)	.890	.879	.884
	P*(H)	.839	.801	.807
	A'	.939	.934	.937
		Group B		
Words Alone	d'	1.69	1.64	1.37
	P(C)	.788	.794	.751
	P*(H)	.651	.723	.632
	A'	.871	.870	.835
Words + Pictures	d'	2.57	2.17	2.83
	P(C)	.879	.834	.901
	P*(H)	.791	.708	.828
	A'	.935	.907	.948

Each measure is based on data combined across Ss.

where P*(H) is the true probability of correct detections, P(H) is the observed hit probability, and P(FA) is the observed false alarm probability. This measure is based on the high threshold assumption that either an item is learned or is not learned during the inspection phase; items which are learned pass the recognition threshold and are correctly recognized; for some proportion of trials on which the threshold is not exceeded, S guesses and his guessing probability is derivable from his false alarm rate (Green & Swets, 1966). A third measure is a nonparametric measure of recognition memory, A' which is a measure of the area of the ROC curve enclosed by a nonparametric ROC curve passing through a single point defined by a pair of hit and false alarm rates (Pollack, 1970; Grier, 1971). A' is calculated according to the following formula:

$$A' = 1/2 + \frac{[P(H) - P(FA)][1 + P(H) - P(FA)]}{[4P(H)][1 - P(FA)]}$$

These four measures are shown in Table 2, for each group and each presentation probability condition. All four measures are approximately the same across the three presentation probability conditions. Furthermore, there is good, although not perfect, agreement among the four measures. The Pearson product moment correlation coefficient across the 12 conditions ranges from a low of +.934 between P*(H) and A' to a high of +.997 between P(C) and A'. P(C) shows the highest average correlation (+.981) with the other three measures.

Table 3 presents averaged values of the four measures of recognition memory for each of the groups. It is apparent that pictorial stimuli produced higher recognition memory scores than words alone. A t test between the words alone and the picture alone stimuli for Group A on the measure P(C) was highly significant (t = 7.31, df = 23, p < .001), as was a t test on P(C) between the words alone and the words + picture stimuli for Group B (t = 4.18, df = 23, p < .001). Furthermore, no improvement in performance was obtained when both the cues, picture and word, were presented over that obtained with the picture alone. There was no significant improvement in recognition memory in terms of differences in P(C) scores between words alone and pictures alone stimuli for Group A and Group B (t = 1.62, df = 46, n.s.). Thus, the presentation of two cues—the word and the picture—did not improve recognition performance over the presentation of a picture alone.

DISCUSSION

The finding that recognition memory is better for pictorial than for verbal stimuli is not surprising. It has been demonstrated repeatedly, both between pictures and words and between high imagery and low imagery words. Presumably, more associative cues may be attached to stimuli which have rich visual as well as verbal associates, so that both recall and recognition performance are improved.

What is surprising, however, is that the addition of verbal and visual cues did not result in superior recognition performance. Certainly, the results of Tulving & Thomson (1971) suggest

that two cues present both during presentation and test should enhance recognition memory performance.

Even on statistical grounds, one would expect an improvement of performance with two cues over one; the exact way in which that improvement is predicted depends upon the particular theory of recognition memory one adopts. Using the concepts of SDT, one predicts that if the two cues are independent, if their memory strengths are normally distributed, and if the standard deviations of the two distributions are assumed equal, the predicted d' from the addition of two signals (or the presentation of two to-be-remembered stimuli, in this case) is given by $d'_{w+p} = \sqrt{{d'_w}^2 + {d'_p}^2}$, where w = words alone, p = pictures alone, and w+p = words + pictures (Green & Swets, 1966). This formula predicts, of course, that the d' measure for both should be higher than for either alone; in particular, using the data from Group A, it predicts that d' for both should be 2.84; using the d' from Group B for words and that from Group A for pictures, it predicts a d' of 2.97.

A high threshold theory of memory, and one which predicts that the two cues are independent and that the S responds positively if either or both cues exceed threshold, predicts the following for P*(H) for w+p:

$$P*(H_{w+p}) = P*(H_w) + P*(H_p) - P*(H_w) P*(H_p).$$

This theory predicts that P*(H_{w+p}) for Group B should be .934 when extrapolated from Group A.

In fact, any model which predicts that both cues are independent and that the S responds to their combination fails to fit the present data; one possible account of the present data is that S stored only the pictorial cue during the words + picture stimuli, and hence responded

Table 3
Average Values for Each Measure of Recognition Memory Across Presentation Probability Conditions for Each Group

		Words Alone	Pictures Alone
Group A	d'	1.31	2.52
	P(C)	.743	.884
	P*(H)	.641	.816
	A'	.828	.927
		Words Alone	Words + Pictures
Group B	d'	1.57	2.52
	P(C)	.778	.871
	P*(H)	.669	.776
	A'	.859	.930

to only that cue during the recognition phase.

REFERENCES

EGAN. J. P. *Recognition memory and the operating characteristic.* Indiana University. Hearing and Communication Laboratory, AFCRC TN 58 51, AD-152650, June 15, 1958.

GREEN. D. M., & SWETS. J. A. *Signal detection theory and psychophysics.* New York: Wiley, 1966.

GRIER. J. B. Nonparametric indexes for sensitivity and bias: Computing formulas. Psychological Bulletin, 1971, 75, 424-429.

MURDOCK, B. B. Signal-detection theory and short-term memory. Journal of Experimental Psychology, 1965, 70. 443-447.

PAIVIO, A. Mental imagery in associative learning and memory. Psychological Review, 1969, 76, 241-261.

POLLACK, I. A nonparametric procedure for evaluation of true and false positives. Behavior Research Methods & Instrumentation, 1970, 2, 155-156.

SHEPARD, R. N. Recognition memory for words, sentences and pictures. Journal of Verbal Learning & Verbal Behavior, 1967, 6, 156-163.

THORNDIKE, E. L., & LORGE, E. *The teacher's word book of 30,000 words.* New York: Bureau of Publications, Teachers College, Columbia University, 1944.

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