Recognition memory as influenced by number of reinforcements and type of test[1]

ROBERTA L. KLATZKY and GEOFFREY R. LOFTUS, Stanford University, Stanford, Calif. 94305

A list of 126 CVCs was presented to each of 29 Ss. Members of the list were reinforced one, two, or three times; and four types of recognition tests were used to evaluate learning of the list. It was found that a signal-detection model applied to the data became increasingly inaccurate as the number of reinforcements of list words increased.

The theory of signal detection (TSD), which treats perceptual processes as statistical decisions, has been applied to recognition memory as well as perception (Bernbach, 1967; Kintsch, 1967; Wickelgren & Norman, 1966). In this application, the presentation of a stimulus stored in memory is viewed as the occurrence of a signal event, while the presentation of an unfamiliar stimulus, or "distractor," is a noise event. The theory assumes that, given a stimulus, the S evaluates his familiarity with it. This value lies on a familiarity continuum; and there correspond to each value on the continuum two probabilities, one of which is the probability that the value occurs if a signal is presented, while the other is the probability of the value given that a noise event occurs. Thus, signal and noise events are associated with signal and noise probability distributions, respectively. It is assumed that both of these distributions are normal; furthermore, the standard deviation of each is set at 1.0 and the mean of the noise distribution is set at zero. The mean of the signal distribution, d', is a parameter estimated from the data which corresponds to the S's sensitivity to previously presented stimuli.

The sensitivity of a S on a recognition task is a measure of his ability to correctly identify words he has studied. The sensitivity parameter, d', should therefore be affected only by those experimental variables which increase or decrease his familiarity with studied stimuli. Events which increase the strength in memory of stimuli presented for study should cause d' to increase, whereas manipulations which do not affect the storage of information about the study words should not change it. In particular, the model predicts that d' will increase as the number of study presentations of stimuli increases. This

should occur because increasing the number of presentations of words on a study list causes the S to become more familiar with those words. In addition, this model predicts that d' will be unaffected by the type of recognition test used. This is expected because the type of test should not affect the S's familiarity with words already studied. The present experiment is designed to test these predictions. It measures the effect on d' of two variables: the number of reinforcements given stimuli on an original study list, and the type of recognition test.

METHOD

The Ss for this experiment were 29 Stanford students, all females, who received $2.00 for an experimental session. The experiment was controlled by a program running on a PDP-1 computer. The Ss were seated at a teletype which displayed the stimuli, and they responded by pressing keys on the teletype keyboard.

The experimental session began with the presentation of a study list of 126 highly pronounceable CVCs (Underwood & Schulz, 1960) at intervals of 2 sec. This list was divided into three "study groups" of 42 stimuli each. Study Group 1 stimuli were presented only once; Study Group 2 stimuli had two presentations; and Study Group 3 stimuli were presented three times. The order of presentation of the CVCs was random, with the restriction that no word was repeated for a second or third time until every word in its group had been presented one or two times, respectively.

Following the presentation of the study list, the S was given 144 test trials. Of these, 36 were yes-no tests, in which Ss responded "yes" or "no" according to whether or not she believed the given word was on the study list. One-half of the yes-no tests used distractors as stimuli. The remaining test trials were forced-choice tests, in which S was shown two, four, or six words, only one of which was presented in the study list; the position of the studied word among the test stimuli was randomly chosen. In such tests, Ss indicated which word was a member of the study list. There were 36 test trials for each type of forced-choice test. Because the type of test was randomly selected at the time it was given, Ss did not know how they would be tested on a word at the time it was presented for study. In addition, confidence ratings of "guess," "moderately sure," and "sure" were given on each test trial.

RESULTS AND DISCUSSION

The probability of a correct response on forced-choice tests and the hit and false-alarm rates on yes-no tests were computed for the group of 29 Ss and are presented in Table 1. (In terms of the theory, a hit is defined as that event in which a studied word is presented and the S correctly identifies it as having been studied. Similarly, a false alarm occurs when a distractor is presented and the S incorrectly identifies it as a member of the original study list.) It can be seen from this table that the hit rate and probability of a correct response on each type of forced-choice test all increase with the number of study-word presentations. Moreover, for words within each study group, the probability of a correct response on forced-choice tests decreases as the number of distractors in the test increases.

The dependent variable of particular interest for this experiment is the parameter d'. As we have stated above, the model generates specific predictions about the effects of the experimental manipulations on the value of d'. First, d' is expected to increase with the number of reinforcements of studied words. Second, we predict that for words in a given study group, d' will remain constant as the type of test is varied. That is, the S's sensitivity should be the same for yes-no tests and forced-choice tests with one, three, or five distractors, as long as all stimuli tested received the same number of presentations during the study period.

In order to test these predictions, the probabilities of Table 1 were converted to d' measures for each test condition and are presented in Table 2. (For a description of the method of converting probabilities to d' measures, see Elliott, 1964.) A two-way analysis of variance was computed over Ss for three numbers of study presentations across four types of tests. The analysis indicates no interaction [$F(6,168) = 2.00$; $p < .10$] and no test type effect [$F(3,84) = 2.24$; $p < .10$], although both of these p-values approach significance. The F

Table 1
Probability of a Correct Response (Forced-Choice Tests) or Hit Rate (Yes-No Tests). The false alarm rate = .32. The theoretical values for these probabilities are in parentheses.

Type of Test	Study Group		
	1	2	3
Yes-No	.57	.75	.77
(Hit Rate)	(.59)	(.73)	(.81)
Forced-Choice,	.70	.73	.83
1 Distractor	(.69)	(.78)	(.83)
Forced-Choice,	.45	.60	.68
3 Distractors	(.46)	(.59)	(.67)
Forced-Choice,	.36	.52	.62
5 Distractors	(.34)	(.47)	(.56)

Table 2
Values of d'

Type of Test	Study Group		
	1	2	3
Yes-No	.65	1.14	1.21
Forced-Choice, 1 Distractor	.74	.86	1.34
Forced-Choice, 3 Distractors	.66	1.12	1.38
Forced-Choice, 5 Distractors	.76	1.23	1.53

for the variable "number of study presentations" is significant, however $[F(2,56) = 50.20; p < .0001]$. Thus, this analysis indicates that d' increases with the number of presentations of words for study, and it does not permit us to reject the hypothesis that d' remains constant as the type of test is varied.

Since no significant change in d' over various methods of test is found for any of the study groups, the mean d' computed over all types of test for words in a given study group should suffice to represent sensitivity to words in that group. This mean d' was used to predict the probability of a correct response on each type of forced-choice test, and, together with the false alarm rate, it was used to predict a hit rate on yes-no tests for words of each study group. Table 1 presents these predicted probabilities. A chi square test was used to determine the goodness-of-fit of the model under the mean d' condition. For this chi square, computed over correct and incorrect responses for all types of tests and all three study groups, the value was 17.12 with 9 df (p = .048). This chi square value indicates that the model does not provide a satisfactory fit to the data. However, its significance is marginal, so that again no strong conclusion can be reached about the constancy of d' over test type.

The ambiguous nature of the results of the two statistical analyses necessitated the use of further tests of the prediction that d' is unaffected by the type of test. To clarify the results of these tests, three one-way analyses of variance were performed. The purpose of these analyses was to test for significant differences in d' corresponding to differences in the type of test, under each of the three reinforcement conditions. The F obtained was not significant for words in Study Groups 1 and 2 $[F(3,112) = .75; p > .50; and F(3,112) = 1.59; p < .25]$; however, a significant F was found for words in Study Group 3 $[F(3,112) = 3.09; P < .05]$. Thus, in the three-reinforcement condition there are significant changes in d' as the type of recognition test varies. Moreover, the values of F obtained by these one-way analyses increase with the number of reinforcements.

These results thus provide an indication that the TSD model begins to break down as the number of reinforcements increases. They suggest that when words are presented once or twice, the mean d' over test conditions is an adequate measure of sensitivity. However, in the case of three reinforcements, it is not possible to represent sensitivity by a single d' value which will hold up over varying types of test.

REFERENCES
BERNBACH, H. A. Decision processes in memory. Psychological Review, 1967, 74, 462-480.
ELLIOTT, P. B. Tables of d'. In J. A. Swets (Ed.), Signal detection and recognition by human observers. New York: Wiley, 1964. Pp. 651-684.
KINTSCH, W. Memory and decision aspects of recognition learning. Psychological Review, 1967, 74, 496-504.
UNDERWOOD, B. J., & SCHULZ, R. W. Meaningfulness and verbal learning. Philadelphia: Lippincott, 1960.
WICKELGREN, W. A., & NORMAN, D. A. Strength models and serial position in short-term recognition memory. Journal of Mathematical Psychology, 1966, 3, 316-347.

Vocalizing choice prior to responding hinders discrimination learning[1]

STEPHEN L. CARMEAN and CHRISTINE R. BAUMAN, Western Washington State College, Bellingham, Wash. 98225

Experiments with both children and adults replicated earlier findings that vocalizing choice before responding resulted in significantly more errors than vocalizing choice after responding. The difference appeared to be due to poorer performance with prior vocalization rather than to better performance with subsequent vocalization.

It was observed in two previous experiments (Carmean, 1969) that Ss, vocalizing their choices in a discrimination learning task, learned more slowly when vocalization was made before, rather than after, responding. This finding was contradictory to the prediction that vocalization before responding should be more effective, if there were any difference at all, because it would cause S to inspect the stimuli more closely before responding.

The two studies reported here were run to replicate this finding and to test two subsidiary hypotheses. One of the previous experiments had used 9-year-old children as Ss, while the other had used adults. In the adult experiment, as in one previously reported (Carmean & Weir, 1967), there was little difference between the group vocalizing after responding and a nonvocal control group, although both made significantly fewer errors than the group vocalizing before responding. In the child experiment, however, the group vocalizing after responding made significantly fewer errors than either the group vocalizing before responding or the nonvocal control group. A nonvocal control group was included in Experiment 1 in an attempt to replicate this finding. Also, since the lack of difference th adults seemed to indicate an interaction with age, even younger children were used in Experiment 1 in order to magnify the effect.

In Experiment 2, an additional condition was added to test delay of knowledge of results as a possible explanation for the poorer performance with prior vocalization. The most obvious difference between the two conditions was that with vocalization before responding, a brief period of time (approximately .7 sec) passed before S knew whether he had vocalized a correct or an incorrect name. Despite the generally accepted assumption (Deese & Hulse, 1967; Hall, 1966) that small delays of knowledge of results are not important with human Ss, it was felt that the possibility that number of errors was a direct function of the amount of elapsed time before feedback was worth checking. This was done by increasing the delay to 4 sec for one group of Ss to see if this would be accompanied by a further increase in errors.

METHOD
Experiment 1
Children in the first grade (mean age = 7.16 years, SD = .51) in Bellingham public schools served as Ss. There were a total of 60 Ss; 20 in each of the three conditions (vocalize choice before responding, vocalize choice after responding, and nonvocal control). Placement in a condition was determined by simple rotation according to order of appearance, with the teachers being allowed to send the children in whatever order was most congruent with the class schedule.

With the exception of the set of stimuli used and the manner of responding, the apparatus and procedure were the same as that previously described (Carmean, in press). Pairs of line drawings of common animals (cow-cat, bear-deer, lion-rabbit, horse-pig, and elephant-giraffe) were presented, one pair at a time, by means of two readout projection units. S made his response by pushing one of two buttons located just below the readout units. The