Recency judgments as measures of temporal coding

ROBERT A. MALMI, ARNOLD M. LUND, and BENTON J. UNDERWOOD
Northwestern University, Evanston, Illinois 60201

The purpose of this study was to determine if recency judgments, used to measure within-list temporal coding, are based on generalized knowledge of temporal relationships among the items in a series or involve learning that is specific to the pairs tested. The subjects studied a 36-item list from which 16 pairs were tested, the subjects making a decision as to which of the two words in each pair occurred most recently in the series. After three trials on which the subjects made recency decisions for the same 16 pairs, the pairings were changed for Trials 4, 5, and 6 for half the subjects. The change in the pairings caused only a small loss. It is concluded that recency judgments made on the same pairings over several trials measure temporal coding that is largely generalizable to any pairing.

A technique used to study changes in intratask temporal coding is that of having the subjects make recency judgments. After a series of discrete stimuli (such as a list of words) is presented, the subjects are shown pairs of stimuli from the series and are asked to make decisions for each pair as to which of the two occurred most recently in the series. This response measure is known to be reliable (Underwood & Malmi, 1978), and the number of correct recency judgments increases across a series of study-test trials. However, when multiple trials are used, a question arises concerning the meaning of the recency judgments when the same pairs of words from the list are used as test pairs on all trials. Specifically, does the improvement in performance across trials indicate that knowledge of the temporal relationships among all items in general is improving, or are the subjects merely learning the most recent item in each pair as a specific association? The present experiment addressed this question.

In the usual experiment using recency judgments, the subjects are instructed before the first study trial as to the method of testing. That is, they are told that they will be asked to decide which is the most recent item in pairs of items taken from the list. Insofar as correct judgments are above chance on the first trial, it indicates that generalized temporal coding has occurred because the subjects did not know how the words would be formed into pairs for testing. This means that recency judgments can be based upon generalized temporal coding and, therefore, might seem to indicate that our concern about specific associative learning was needless. There are two considerations which moderate this

matter. First, performance on the first test trial is usually low, even if better than chance. For example, in the study cited above, there were 63% correct recency judgments on the first test trial. Second, it may be that on subsequent trials the subjects more and more turn to specific associative learning so that there may be no more generality to the knowledge of temporal coding on later trials than on the first. In any event, it was our intent to discover the degree of generality of temporal coding resulting from multiple trials in which the test pairs were the same on all trials.

METHOD

All subjects received six study-test trials. Half of the subjects had the same test pairs on all six trials, and half had the pairs changed between the third and fourth trials. If the performance of the second group were to fall sharply on the fourth test trial, it would be taken as evidence that at least some of the recency judgments were based on learning that was specific to the pairs.

List and Test Pairs

The list consisted of 36 AA words (Thorndike & Lorge, 1944), all of five letters. The 36 words were ordered randomly, and this order was used for all study trials for all conditions. The words holding Positions 1, 2, 35, and 36 were never tested. The remaining 32 words were paired randomly, and this pairing will be called P1. The mean lag (number of other items falling between two test words in the study list) was 11.13. A second pairing (P2) was made up so that the mean lag was also 11.13, and the standard deviation of the lags was about the same as that for P1. The same two words were never used as a test pair in both P1 and P2.

Procedure and Subjects

To balance potential differences in difficulty between P1 and P2, it was necessary to use four groups of subjects. Two of these groups (control groups) had all six trials on the same pairings, one group having P1, the other P2. Another group had the first three trials on P1 followed by three trials on P2, and a fourth group had the reverse. The four groups may then be identified as P1-P1, P2-P2, P1-P2, and P2-P1, with three trials on each side of the hyphen. A total of 20 college students was assigned to

Table 1
Mean Correct Recency Judgments (16 Possible)
for Each Condition on Each Trial

Trial	Condition			
	P1-P1	P2-P2	P1-P2	P2-P1
1	10.00	11.00	10.95	10.80
2	12.40	12.55	12.45	11.90
3	12.95	13.10	12.70	13.25
Mean	11.78	12.22	12.03	11.98
4	14.25	13.95	13.00	13.65
5	14.50	14.65	13.85	14.20
6	14.75	15.10	13.75	14.30
Mean	14.50	14.57	13.53	14.05

each of the four groups by a block-randomized schedule.

The study list was presented at a 4-sec rate, and the recency tests were paced at a 6-sec rate. The subjects were fully instructed about recency judgments before the first study trial. Groups P1-P2 and P2-P1 were not told that the pairings would change on the fourth test trial.

RESULTS AND DISCUSSION

The mean numbers of correct recency judgments for each trial are shown in Table 1. The means for Trials 1, 2, and 3 combined and for Trials 4, 5, and 6 combined are also given. A comparison of the scores on the first three trials indicates whether or not the two sets of pairings (P1 and P2) produced lists of different difficulty, or whether the experimental and control groups differed in ability levels. Using P1 and P2 as one variable and experimental-control as another, the analysis of variance did not produce an F larger than 1. The four groups must be considered statistically equivalent on the first three trials.

The critical data are the scores on Trials 4, 5, and 6. An examination of Table 1 shows that the performance of the two experimental groups (P1-P2; P2-P1) did not decrease between Trials 3 and 4 as would be anticipated if recency decisions were based on associations specific to each pair. However, a careful perusal of the scores on Trials 3 and 4 shows that the increase between these two trials was less for the two experimental groups than for the two control groups. A difference score (Trial 4 minus Trial 3) was determined for each subject, and an analysis of variance was performed. The mean difference scores were not statistically larger for the control groups than for the experimental groups [$F(1,76) = 2.84$, $p > .05$].

Table 1 shows that the scores of the experimental group remained below those for the control group on Trials 5 and 6. An analysis for the scores summed across Trials 4, 5, and 6 showed that the performance of the control group was statistically better than that for the experimental groups [$F(1,76) = 5.53$, $p < .05$]. Roughly speaking, the data indicate that switching the pairings between Trials 3 and 4 caused the subjects to lose one correct recency judgment, and this loss was not made up in the remaining two trials. It seems proper to conclude that recency judgments are primarily measuring generalized temporal information. The amount of the learning that is specific to the test pairs is minimal.

REFERENCES

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

UNDERWOOD, B. J., & MALMI, R. A. An evaluation of measures used in studying temporal codes for words within a list. *Journal of Verbal Learning and Verbal Behavior*, 1978, in press.