

Effects of the number of study trials on single-item recognition following verbal discrimination learning

ROBERT W. NEWBY

Tarleton State University, Stephenville, Texas 76401

Sixty subjects were exposed to one of four 20-item verbal discrimination lists for one, two, or four study trials. A recognition task followed consisting of the previous correct and incorrect items and a set of distractor words. The results of the analysis of the recognition task indicated superior discriminability of correct items as compared to incorrect items. The results also suggested that one criterion had been adopted which appeared to be lenient for correct items and stringent for incorrect items. The results were interpreted in terms of frequency theory.

Underwood, Jesse, and Ekstrand (1964) studied retention of items across lists by means of a verbal discrimination transfer task. Two transfer situations were investigated: lists containing the same correct items and lists using the same incorrect items. Performance was essentially perfect when the first-list correct items were retained, whereas performance was good but showed no improvement with retention of first-list incorrect items which indicated recognition of first-list items. Underwood et al. explained these results by assuming that second-list discrimination performance is based on the relative frequency of the two items. The older unit has more frequency than the new unit and is therefore more recognizable.

Based on these results, Ekstrand, Wallace, and Underwood (1966) proposed the frequency theory of verbal discrimination learning, which states that during learning responses are made to the correct and incorrect members of a verbal discrimination list that adds frequency units to these items. According to the theory, the correct item obtains more frequency units than the corresponding incorrect item. Assuming that recognition performance is based on frequency differences, this should make the correct item more recognizable. This assumption was tested by Erlebacher, Hill, and Wallace (1967) by giving subjects a recognition test following learning to a criterion of one perfect trial. The results indicated a better recognition of correct than incorrect items, which corresponds to the prediction based on the frequency theory.

Radtke and Foxman (1969) extended this finding by comparing recognition performance following two-alternative and four-alternative verbal discrimination

learning. Frequency theory predicts that correct items would be more easily recognized following four-alternative learning than in the two-alternative situation. Recognition performance of incorrect items should show the opposite result. Radtke and Foxman found that recognition of correct items was essentially perfect in both conditions, while recognition of incorrect items was superior in the two-alternative situation. The authors interpreted these results as supportive of frequency theory.

Recently, Kausler, Pavur, and Yadrick (1975) applied signal detection theory to verbal discrimination learning. This method allows the experimenter to determine the criterion used by subjects in old-new recognition. Following one study trial, the subjects' hit rates for previous correct and incorrect items were analyzed. They found higher hit rates for correct than for incorrect items, supporting the prediction from frequency theory. These results indicate that, following one trial of verbal discrimination learning, there are three theoretical distributions based on frequency: one for old correct items, one for old incorrect items, and one for new items. In addition, Kausler et al. suggest that the subject uses a relatively lenient criterion for old-new item recognition.

Frequency theory predicts that on each study interval correct items should acquire more frequency units than the incorrect items. In situations involving multiple study trials, this should serve to increase the difference between the three distributions proposed by Kausler et al. (1975). Thus, as study intervals increase, recognition of correct items should proceed at a faster rate than recognition of incorrect items.

To test this assumption, three study intervals were selected (one study trial, two study trials, and four study trials). Following learning in all conditions, it was assumed that recognition of correct items would be greater than that of incorrect items and that this difference should increase as the number of study trials increased. Increasing the study time should augment the

The author would like to express his thanks to Robert E. Gehring and Michael P. Toglia for their aid in the preparation of the data analysis. This research was funded by an organized research grant from Tarleton State University. Robert K. Young sponsors this paper and takes full editorial responsibility for its contents.

discriminability of correct items compared to incorrect items.

METHOD

Design

The three amounts of study time were factorially combined with two lists and two sublists to give 12 experimental groups. Included as a within-subjects variable was item function, correct vs incorrect. There were five subjects in each experimental group.

Procedure and Apparatus

Two groups of 40 words were selected from the Paivio, Yuille, and Madigan (1968) norms, keeping in mind that the items should be of equal imagery and frequency. Each list was divided randomly into two sublists, A and B. These sublists served as correct and incorrect items for the verbal discrimination lists. The sublists of the other group of words served as the distractor items during recognition. Therefore, if items in Sublist 1A were correct and items in 1B were incorrect, then Sublist 2A contained distractors for the correct items, while 2B served the same function for incorrect items. All words served each of the four functions—correct items, incorrect items, distractors for correct items, and distractors for incorrect items—making four 20-item verbal discrimination lists.

During the study intervals all pairs of the verbal discrimination lists were presented at a 3-sec rate by a Kodak 850H Carousel projector. With multiple study trials, there was a 12-sec intertrial interval. Timing was done manually by means of a stopwatch. All pairs, with the correct item underlined, were presented on 5 x 5 mm slides. The position of the correct item (top or bottom) was randomly varied to control for position effects. In the two and four study trial conditions, the order of presentation was varied to control for serial order effects.

All subjects received the same recognition task. Eighty items were presented, one at a time, each for 3 sec. For each item, the subject was required to mark on a scale from 1 to 7 whether the item was "old" or "new." Seven represented the subject's certainty of "oldness," while 1 represented certainty of "newness." Again for each group of subjects, 20 items were old and correct, 20 were old and incorrect, and 40 were distractors (20 for the correct items and 20 for the incorrect items). These items had also been typed individually on 5 x 5 mm slides and the four types of words were presented in a random sequence.

Subjects

The subjects in the present study were 60 introductory psychology students at Tarleton State University. Each subject received extra credit for his/her participation. All were naive with regard to verbal discrimination and recognition tasks.

RESULTS AND DISCUSSION

As a measure of discriminability, areas (A) under the ROC curves (Banks, 1970) for both correct and incorrect items were calculated for each subject. This was chosen because it is a nonparametric measure of performance and it is distribution free. An analysis of variance was calculated on the scores. The analysis indicated a significant effect of study interval [$F(2,48) = 12.67, p < .001, MSe = .024$]. As the number of study trials increased, so did the discriminability of the "old" items. The mean A scores for the three groups were .57, .60, and .72 for the one study trial, two study trial, and

four study trial groups, respectively. Also, a significant item function effect was found [$F(1,48) = 62.00, p < .001, MSe = .011$]. The mean A value for correct items was .70, while the mean value for incorrect items was .55. No other main effects or interactions were found to be significant.

To calculate the criterion measures, the hit and false alarm rates were calculated for each subject. This was done by collapsing the confidence rating scale into a yes-no scale. Thus, old items receiving a score of 5 or greater were considered hits, while distractors receiving the same score were considered false alarms. Criterion scores were then calculated for correct and incorrect items using the percent bias formulas described by Hodos (1970). With this scale, scores may range from +100% to -100%, with a positive number indicating a strict criterion and a negative number indicating a lenient criterion. The mean bias scores for correct and incorrect items for each study interval are presented in Table 1. The main effect for item function was significant [$F(1,48) = 29.25, p < .001, MSe = 1,167.37$]. This seems to indicate that a stricter criterion was set for responding to incorrect items than to correct items. This varied according to study interval, as indicated by a significant interaction of Item Function by Study Time [$F(2,48) = 4.04, p < .05, MSe = 1,167.37$]. The criterion set for responding to correct items became more lenient with increased study time, while the criterion used for incorrect items became stricter.

The significance of these results could possibly be diminished by the following group of significant effects and interactions involving the sublists. In the analysis of criterion bias scores, there was a significant sublist effect [$F(1,48) = 6.82, p < .05, MSe = 2,377.84$] which was due to a stricter criterion for correct items from the B sublists than those from the A sublists. Also significant was the interaction of Item Function by Lists by Sublists [$F(1,48) = 4.88, p < .05, MSe = 1,161.37$]. This interaction resulted from the criterion set for items making up the A sublists in the two main lists. A more lenient criterion was set for 1A items compared to 2A items when these served as incorrect members, while the situation was reversed when the A items were correct members. A significant four-way interaction between all variables [$F(2,48) = 4.61, p < .05, MSe = 1,161.37$] was

Table 1
Mean Bias Percentages at Each Study Interval

Study Interval	Item Function	
	Correct	Incorrect
1 Trial	8%	20%
2 Trials	-14%	21%
4 Trials	-12%	43%
Mean Bias for Item Function	- 6%	28%

also found. This interaction was due to differences in criterion for correct and incorrect items set by subjects in each of the experimental groups. Although each was significant, these effects do not detract from the main finding of two criteria by which to respond to correct and incorrect members as old items.

Two counterintuitive points emerge from the analyses of the data collected in this experiment. First, based on the results of the analysis of A scores, the discriminability between correct and incorrect items does not increase over study trials. This should occur based on predictions derived from frequency theory. Also, the results suggest that the subjects have adopted two different criteria to differentiate between old and new items, one for previously correct items and the second for previously incorrect items. McFadden (Note 1) states that it is far more likely that there is but one criterion used to judge oldness.

To clear some of the ambiguity, two additional analyses were performed on the hit and false alarm data. If, as the results suggest, the discriminability between correct and incorrect items does not increase over trials, then correct and incorrect items should show similar increases in hit rates and decreases in false alarm rates. In addition, if subjects do use two criteria to determine oldness, there should be differences between correct and incorrect items in the false alarm analysis.

The analysis of hit data revealed two significant effects. Hit rates increased with increasing study time [$F(2,48) = 6.06, p < .01, MSe = .036$] and correct items were more often correctly identified as "old" than were incorrect items [$F(1,48) = 99.97, p < .001, MSe = .029$]. The analysis of false alarms revealed a significant effect for study trials [$F(2,48) = 3.50, p < .05, MSe = .034$] which indicated that the false alarm rate decreased with increasing study time. There was no significant item function effect, indicating no differences in false alarms between correct and incorrect items. There were additional significant interactions in the false alarm analysis at probability level $p < .01$. These were Item Function by List [$F(1,48) = 9.00, MSe = .005$], Item Function by List by Sublist [$F(1,48) = 19.00, MSe = .005$], and Item Function by List by Sublist by Study Time [$F(2,48) = 8.40, MSe = .005$]. These interactions were due to different frequency distributions among the four groups of verbal items (1A, 1B, 2A, and 2B) and do not alter the interpretation of the other main effects and interactions.

A summary of all analyses, then, suggests that following verbal discrimination learning there are three distributions of verbal materials: previously correct items, previously incorrect items, and distractors. This finding replicates that of Kausler et al. (1975). The discriminability between old and new items increases with increasing study time. However, the difference between correct and incorrect items remains relatively constant, a finding not completely supportive of frequency theory.

The results also indicate that the subjects adopt a single criterion by which to judge "oldness." This idea was supported by the false alarm analysis, which found no difference between previously correct and incorrect items. By its placement on the frequency continuum, a greater proportion of correct items lie above the criterion than do incorrect items. This gives the appearance of strict and lenient criteria, when in fact only one is being used. This finding also replicates Kausler et al.

REFERENCE NOTE

1. McFadden, D. Personal communication, 1976.

REFERENCES

- BANKS, W. P. Signal detection theory and human memory. *Psychological Bulletin*, 1970, **74**, 81-99.
- EKSTRAND, B. R., WALLACE, W. P., & UNDERWOOD, B. J. A frequency theory of verbal discrimination learning. *Psychological Review*, 1966, **73**, 566-578.
- ERLEBACHER, A., HILL, W. J., & WALLACE, W. P. Differential accrual of frequency in verbal discrimination learning. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 420-422.
- HODOS, W. Nonparametric index of response bias for use in detection and recognition experiments. *Psychological Bulletin*, 1970, **74**, 351-354.
- KAUSLER, D. H., PAVUR, E. J., & YADRICK, R. M. Single-item recognition following a verbal discrimination study trial. *Memory & Cognition*, 1975, **3**, 135-139.
- PAIVIO, A., YUILLE, J. C., & MADIGAN, S. Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monograph Supplement*, 1968, **76**(1, Part 2).
- RADTKE, R. C., & FOXMAN, J. Number of alternatives and similarity in verbal discrimination learning. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 537-544.
- UNDERWOOD, B. J., JESSE, F., & EKSTRAND, B. R. Knowledge of rights and wrongs in verbal discrimination learning. *Journal of Verbal Learning and Verbal Behavior*, 1964, **3**, 187-194.

(Received for publication December 19, 1976.)