Similarity in short-term recall: A comparison of Type II operating characteristics and percent recalled

ROBERT WEEKS and LEONARD KATZ
University of Connecticut, Storrs, Conn. 06268

The effects of semantic and acoustic similarity in short-term memory were examined in an experiment employing an RI paradigm. In assessing similarity effects, signal-detection measures, Type II operating characteristics, were utilized in addition to percent recalled; the former suggested that both semantic and acoustic similarity between OL and IL were important in affecting recall, while the latter revealed no effects of similarity. The results suggested that memory traces for similar and dissimilar OL items were equal in strength but that noise (due to RI) was stronger for the similar conditions.

The importance of acoustic similarity in short-term memory (STM) and semantic similarity in long-term memory (LTM) has been demonstrated in many experiments. In an RI paradigm where similarity between IL has been varied, the effect of acoustic similarity on STM (Dale, 1964; Wickelgren, 1965) and of semantic similarity on LTM (Robinson, 1927; Baddeley & Dale, 1966) have been found.

The effects of similarity on recall have traditionally been shown by the measures of percent recalled, by intrusions, and by omissions. However, Murdock (1966) has demonstrated that the traditional measures utilized in short-term memory experiments may overlook the effects of changing response criteria. Such changes may reflect changing strategies of Ss with changes in experimental tasks or material (cf. Katz, 1970).

The present experiment compares the traditional measure of percent recalled with a measure from the theory of signal detection (TSD) in a short-term memory experiment. An RI paradigm was employed with the acoustic or semantic similarity between OL and IL varied.

METHOD

The four conditions run in a S by Treatments design were semantic-dissimilar (SemD), semantic-similar (SemS), acoustic-dissimilar (AcD), and acoustic-similar (AcS). For each condition, 12 sets of six monosyllabic words were employed. For example, one set in the acoustic conditions was: "wraps, traps, claps, gaps, flaps, slaps." On each AcS trial, three words from one set, e.g., "slaps, wraps, flaps," were presented together on a screen to the Ss for 2.0 sec (OL). Then a second slide was immediately shown for 2.0 sec (IL); this contained the three remaining words from the set, e.g., "claps, gaps, traps." In AcD trials, OL consisted of three words from another acoustic set, e.g., "come, plum, crumb." The semantic-condition sets were composed of the first six monosyllabic words in categories from the category norms of Battig & Montague (1969), and for the semantic conditions, experimental material was constructed and temporal design proceeded analogously to that of the acoustic conditions.

In all conditions, immediately after the presentation of the IL words, the Ss shadowed and summed pairs of digits (presented orally by E at the rate of one pair/second) in data booklets for 4.0 sec. The Ss then attempted to recall the IL words in any order for 10.0 sec, at which time Ss were signaled to recall the OL words for 15.0 sec. The Ss were instructed that they must give three responses for each word triad that they were attempting to recall and must attach to each response a confidence rating, R_j, $j = 0, 1, \cdots, 5$, where a rating of 0 meant "Positive I recalled the word incorrectly" and a rating of 5 meant "Positive I recalled the word correctly." Since Pollack & Decker (1958) found that the production of confidence ratings did not impair accuracy of message reception, it seemed likely that such ratings would not interfere with recall in the present experiment.

Each S was presented 12 trials in each of four conditions. The order of conditions was counterbalanced among groups to produce four orders. To control for any bias that could result from the use of particular OL or IL material, two groups were employed in each order, with the OL and IL of one group reversed for the other group in the SemS and AcS conditions. The same OL in the SemD and AcD conditions was used for all groups, but within groups, the IL in the dissimilar conditions was identical to the IL of the similar conditions.

All Ss were from the introductory psychology course at the University of Connecticut. Each of the eight groups contained from 11 to 16 Ss.

RESULTS AND DISCUSSION

All correct response frequencies within each condition in each group were converted to percentages, and the mean percentage across groups was computed for each condition; thus, all groups contributed equally to the data analysis. Table 1 shows the mean percentages correctly recalled and the respective standard deviations of percentages for the eight groups of each condition.

The percent-recalled measure does not reveal any effect of either semantic or acoustic similarity on the recall of OL. The reason for the departure of the present finding, employing the percent-recalled measure, from the more frequent finding (e.g., Baddeley & Dale, 1966), that acoustic similarity between OL and IL impairs recall of OL, is unknown. However, there are several differences between the present experiment and previous ones, viz, the requirement that the Ss recall IL as well as OL and the requirement that the Ss record three responses with certainty judgments recalling OL on each trial, that may have produced the result.

The signal-detection-theory analysis indicated that both semantic and acoustic similarity produced decrements in the strength of a recall response. The Type II (i.e., response conditional) operating characteristic curves of Fig. 1 show that similarity between OL and IL decreases the relative strength of an elicited response. The d′ values in Table 1 show that similarity decreases d′ by about 0.8. Thus there is an apparent discrepancy between the percent recall measure and the TSD measure.

This discrepancy may, however, be explained if signal strength of OL is considered to be equal for both similar and dissimilar conditions. Percent correctly recalled for both conditions should be equal if Ss are forced to always respond. But if the means of the noise strength distributions were greater in the two similar conditions, any given criterion, i.e., liberal ($R_j = 0$), would yield higher error rates in the similar conditions that would result in lesser d′ values. In the similar conditions, Ss were less likely to express

Table 1
The Mean and the Standard Deviation of Percent Recalled Correctly and the d′ for Each Condition

Condition	Percent Correctly Recalled	SD Percent Correctly Recalled	d′
SemD	60.86	9.19	2.30
SemS	56.69	7.61	1.47
AcD	34.11	6.48	1.96
AcS	35.44	5.72	1.22

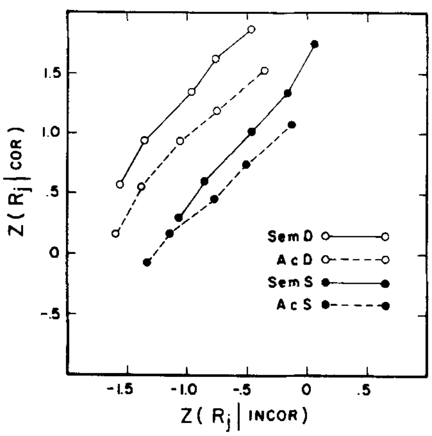

Fig. 1. The receiver operator characteristic (Type II) averaged over all groups for the four OL-IL similarity conditions. Each point is based upon a number of responses that ranges between 156 and 1,656.

strong confidence in either correctness or incorrectness of responses. For example, the probability that an S uses a 5 rating (very confident response is correct) when actually incorrect, P(5 | INCOR), is slightly greater in the similar conditions. P(5 | INCOR) was: SemD, 0.059; SemS, 0.148; AcD, 0.055; AcS, 0.091. Nevertheless, there is some evidence that Ss' criterion placement for the low-confidence categories changed from the similar conditions to the dissimilar conditions. For example, P(0 | INCOR) is lower in the similar conditions, e.g., SemD, 0.680, SemS, 0.475, AcD, 0.676, AcS, 0.543, data that are not compatible with the fixed-criteria notion since noise should overlap more with the signal distribution in the similar conditions and thus produce more errors in that condition. This suggests that an S adopts a generally more conservative strategy in evaluating responses (at least for the $R_j = 0$ decision) perhaps because S realizes the difficulty of the similar material, i.e., he is sensitive to the reduced d′.

The present results suggest that Es, who would employ traditional measures in experiments where Ss are not forced to respond, are risking the confounding influences of the criterion problem to which Murdock (1966) refers. It seems quite probable that many experimental manipulations do not passively affect Ss' behavior but initiate the active shifting of response criteria by Ss who are adopting strategies toward the experimental tasks before them.

REFERENCES

BADDELEY, A. D., & DALE, H. C. A. The effect of semantic similarity on retroactive interference in long- and short-term memory. Journal of Verbal Learning & Verbal Behavior, 1966, 5, 417-420.

BATTIG, W. F., & MONTAGUE, W. E. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut Category Norms. Journal of Experimental Psychology, 1969, 80 (3, Part 2).

DALE, H. C. A. Retroactive interference in short-term memory. Nature, 1964, 203, 1408.

KATZ, L. A comparison of Type II operating characteristics derived from confidence ratings and from latencies. Perception & Psychophysics, 1970, 8, 65-68.

MURDOCK, B. B., JR. The criterion problem in short-term memory. Journal of Experimental Psychology, 1966, 72, 317-324.

POLLACK, I., & DECKER, L. R. Confidence ratings, message reception, and the receiver operating characteristic. Journal of the Acoustical Society of America, 1958, 30, 286-292.

ROBINSON, E. S. The "similarity" factor in retroaction. American Journal of Psychology, 1927, 39, 297-312.

WICKELGREN, W. A. Acoustic similarity and retroactive interference in STM. Journal of Verbal Learning & Verbal Behavior, 1965, 4, 53-61.

The effects of state instructions and schedules of reinforcement on resistance to extinction*

RONALD K. PARKER†
Florida State University, Tallahassee, Fla. 32306

College females (N = 70) were used in an experimental design manipulating state instructions and seven schedules of reinforcement (0% to 100% in increments of 1/6). The state instructions required the Ss to guess before each trial if they were ever going to be reinforced again throughout 30 acquisition trials and 60 extinction trials—i.e., the S guessed whether she was in an experimental state of acquisition or extinction. The most lean schedules of reinforcement (1/6 and 2/6) were most resistant to extinction; however, the resistance to extinction was not related in an orderly fashion to schedule of reinforcement.

Parker (1967) distinguished between the typical "trial instructions" (TI) employed in probability learning experiments in which the S guessed before each trial if he was to be rewarded on the *next trial* (e.g., Grant, Hake, & Hornseth, 1951) and "state instructions" (SI) in which the S guessed if he would ever be rewarded again throughout the remainder of the experimental session—i.e., the S guessed whether or not he was in an experimental *state* of acquisition or extinction. Using children as Ss, he manipulated instructions, schedules of reward, and magnitude of reward in a 2 by 2 by 2 factorial design. The "state" and "trial" instructions produced distinctly different behavior in acquisition and extinction, and, using 50% and 100% schedules of reward, the partial reinforcement effect (PRE) was more clearly evidenced in SI than in the TI.

This experiment represents one in a series of studies designed to explore the

*This research was supported by a Florida State University Research Council award, PARK 35-036.

†Now at the Center for Advanced Study in Education, The City University of New York.