Journal of
Experimental Psychology

VOL. 70, NO. 5 NOVEMBER 1965

SIGNAL-DETECTION THEORY AND SHORT-TERM MEMORY [1]

BENNET B. MURDOCK, JR.[2]

University of Vermont

This study attempted to apply signal-detection theory to short-term memory by testing the high-threshold concept of associative strength. On each trial a list of 6 A-B pairs was presented once, then 1 of these 6 pairs was tested for recognition. On the recognition test either A-B (a proper pair) or A-X (an improper pair) was presented; S had to make a binary (yes-no) decision plus a confidence rating. From these data ROC curves were plotted, and they resembled the curvilinear functions of signal-detection theory more than the linear function required by the high-threshold concept. These results call into question the use of a high-threshold concept to explain findings from studies of 1-trial learning.

Recent developments in signal-detection theory have provided a "new look" at many problems in perception (Swets, Tanner, & Birdsall, 1961) and attention (Broadbent & Gregory, 1963). Basically, signal-detection theory makes possible a separation of the criterion and the sensitivity of *O* and provides some powerful tools for the analysis of these processes. It seemed likely that these methods could profitably be applied to problems in short-term memory, and the present study is such an attempt.

Specifically, the present study is an attempt to test the "high-threshold" concept in short-term memory. Underwood and Keppel (1962), in an article on one-trial learning, have suggested a threshold of associative strength below which no response will occur. Thus, results from studies which use an RTT paradigm can be explained; items above the threshold on T_1 will tend to be above the threshold on T_2 while items below the threshold on T_1 will tend to be below the threshold on T_2. However, as Estes (1962) has pointed out, recent threshold studies do not provide much support for this strong concept. The purpose of the present experiment was to try to make a direct test

[1] This investigation was supported by Public Health Service Research Grant MH 03,330 from the National Institute of Mental Health, Public Health Service. The writer is endebted to Donald A. Norman for helpful comments and suggestions. Cynthia M. Fischer assisted in collecting and analyzing the data.

[2] Now at the University of Toronto.

of the high-threshold concept in short-term memory itself.

The procedure used in the present experiment followed directly along the lines of signal-detection theory. Short-term memory was tested by a recognition test supplemented with confidence ratings. As in previous studies (e.g., Murdock, 1963) on each trial a list of six paired associates was presented, then retention of one of the six pairs was tested. Here, however, retention was tested by recognition; *S* was shown either A-B (i.e., one of the pairs from the list) or A-X (i.e., not one of the pairs from the list) and asked to give both a yes-no judgment and a confidence rating. As has been pointed out (e.g., Egan, Schulman, & Greenberg, 1959; Pollack & Decker, 1958), with confidence ratings one has the necessary information to plot ROC curves. The question of interest, then, is whether the resulting ROC curves are the linear function demanded by the high-threshold concept (Swets, Tanner, & Birdsall, 1961).

Procedure

All *S*s were tested individually and were given 54 trials on each of four sessions. Each trial consisted of a presentation of six A-B pairs followed by a recognition test for one of the six pairs. The recognition test consisted of presenting either an A-B or an A-X pair; *S* had to respond either "yes" or "no" and, in addition, give a confidence judgment. An A-B pair would be a proper pair while an A-X pair would be an improper pair; that is, two words which had and had not been paired during list presentation. Correct responses were, of course, yes and no, respectively.

The serial position of the critical pair (i.e., that pair tested in the recognition test) was defined in terms of the first (or A) member of the A-B pair. With 54 lists on four sessions there was a total of 216 trials, and each of the six serials positions (SP) was tested equally often. For each serial position half the recognition tests were A-B ("signal") while half were A-X ("noise"). For Serial Positions 1 and 6 the X terms were always the second (or B) members of the pairs from Serial Positions 2 and 5, respectively. For Serial Positions 2–5 the X terms came equally often from the immediately preceding and succeeding pair (though always the B member of the adjacent pair).

All pairs consisted of common English words taken from the Thorndike-Lorge (1944) list with G count of 20 or above. All homonyms, contractions, archaic words, and words with more than eight letters were excluded. Words were paired at random and without restriction. Pairs were photographed on 16-mm. film and projected by a single-frame Dunning animatic projector. The two words in each pair were displayed vertically rather than horizontally (i.e., one above the other rather than side by side). A visual noise field was inserted between each pair; this visual noise field consisted of letters of the alphabet typed at random and as densely as possible. Thus, on the filmstrips A-B pairs and the visual noise field alternated regularly in an ababa order.

There were four different filmstrips, one for each session, and their order of presentation was counterbalanced across *S*s in a 4×4 Latin square. Also, orthogonal to this counterbalancing A-B and A-X pairs were reversed for half the *S*s. That is, any pair which had been tested as A-B for half the *S*s was tested as A-X for the other half, and vice versa. Within each of the filmstrips the order of presentation of conditions was randomized in blocks of 12 trials (i.e., six serial positions and A-B or A-X test trials), and this randomized order naturally differed on the four filmstrips.

Stimulus presentation was programed by Hunter timers; A-B pairs and visual noise fields alternated at a 1-sec. rate. Each list presentation began and ended with a visual noise field. At the end of each list (i.e., when the visual noise field following the sixth pair had appeared) *S* was handed an index card with two words written on it, one above the other. He then wrote down on his data sheet either "yes" or "no"; yes if it matched one of the pairs in the list and no if it did not. Also, he had to write down a number from 1 to 5 indicating how confident he was that he was correct. In the instructions *S* was told, "1 is a pure guess, 3 is average certainty, 5 if you are very sure, and 2 and 4 are degrees between to help you rate." Also, a description of this rating scale was constantly in view for *S* to refer to throughout the experiment.

There were 24 *S*s, 12 students of each sex who were fulfilling a requirement of the introductory psychology course. On the first day there were two practice trials, but none thereafter. After each list presentation there was a 15-sec. interval for *S* to respond, and he was instructed to make a choice on every trial. The end of the 15-sec. interval was indicated by an audible click from one of the timers, and a "ready" signal preceded the start of the next list by several seconds.

RESULTS

The probability of a correct recognition (i.e., yes for A-B and no for A-X) was .764, .745, .720, .730, .874, and .927 for Serial Positions 1–6, respectively. An analysis of variance (of the number of correct recognitions) showed that the effects of serial position were highly significant; $F(5, 115) = 27.55$, $p < .001$. However, individual *t* tests showed that, at the .05 level, none of the comparisons involving the first four serial positions differed significantly from one another. Therefore, SP 1–4 were combined for subsequent analysis.

The ROC curves for SP 1–4 and for SP 5 are shown in Fig. 1. The curves were constructed following the method described by Pollack and Decker (1958). To indicate this method, the raw data for SP 1–4 are shown in Table 1. With four serial positions, 18 replications, and 24 *S*s, there were a total of 1,728 trials for the A-B and for the A-X presentations. The first two rows of

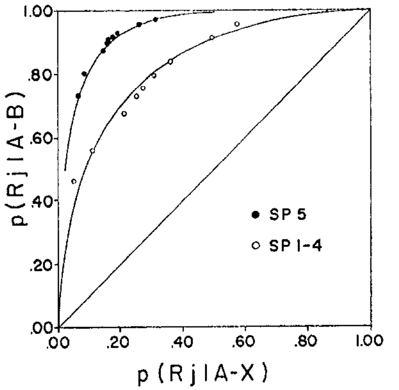

FIG. 1. Cumulative conditional probabilities of ratings R_j or stricter for A-B and A-X presentations.

Table 1 show the distributions of these 1,728 trials over the 10 possible confidence judgments. The confidence judgments are ordered left to right from "No-5" to "Yes-5" to correspond with a criterion ordering from lax to strict, respectively. The conditional probabilities, shown in the bottom two rows of Table 1, are cumulative, the probability that a given confidence rating was assigned to Category *j or stricter*. Thus, one starts from the right and moves to the left, separately for A-B and A-X.

The typical ROC curve shows hits (ordinate) as a function of false alarms (abscissa). Here, the cumulative

TABLE 1

DISTRIBUTION OF CONFIDENCE JUDGMENTS AND CONDITIONAL PROBABILITIES FOR A-B AND A-X PRESENTATIONS (SUMMED OVER FOUR SERIAL POSITIONS, 18 REPLICATIONS, AND 20 *S*S)

	No					Yes					Total
	5	4	3	2	1	1	2	3	4	5	
A-B	75	75	128	80	65	49	93	203	165	795	1728
A-X	732	137	231	94	58	37	73	172	102	92	1728
$P(R_j \mid \text{A-B})$	.957	.913	.839	.793	.755	.727	.673	.556	.460		
$P(R_j \mid \text{A-X})$	.576	.497	.363	.309	.275	.254	.212	.112	.053		

conditional probability of assignment to Category j or stricter is shown for A-B presentations (ordinate) as a function of A-X presentations (abscissa). The rationale behind this method has been given by Egan (1958); Egan, Schulman, and Greenberg (1959); and Pollack and Decker (1958). Basically, the idea is that *S* is capable of adapting multiple criteria, and a rating method provides more information about these criteria than one obtains from a binary (yes-no) judgment. In the present case, a judgment of "yes-5" represents the strictest criterion; "yes-4" is next most strict, and any observation that exceeded the criterion in the former case would also do so in the latter. Then, with 10 ratings, there are nine criteria ranging, as indicated, from strict to lax.

The data points for SP 1–4 and for SP 5 are well fitted by the smooth curves of signal-detection theory. (Data for SP 6 are not shown because the confidence ratings were so extreme.) In order to construct these smooth curves the data points were first plotted on normal-normal coordinates (Swets, Tanner, & Birdsall, 1961) and the best-fitting straight line drawn by eye. For SP 1–4 the slope of this line was close enough to unity to accept the assumption of equal variances for the A-B and the A-X distributions, but for SP 5 the slope was approximately 1.25. The obtained values for d' were 1.36 and 2.0 for SP 1–4 and SP 5, respectively. The difference in symmetry between the two curves simply reflects the fact that the equal-variance assumption held only for SP 1–4.

As has been pointed out elsewhere (e.g., Egan, 1958; Swets, 1961), the high-threshold assumption predicts that the data points should lie along a straight line running from the left-hand vertical axis to the upper right-hand corner of the graph. The y intercept would indicate the "true" threshold and the slope would be equal to one minus this threshold proportion. The linear function is a consequence of the fact that, according to a high-threshold assumption, *O* is incapable of ordering his judgments below the threshold. Thus, an increase in the hit rate can only be achieved by more false alarms, and these must increase at the same rate. As can be seen from Fig. 1, the data points do not fall along a straight line, and it would seem that the results do not support the high-threshold assumption.

Discussion

From the results of the present experiment it does not seem correct to postulate a high threshold in short-term memory (STM), above which one can make correct recognitions and below which one responds by chance. Instead, recognition in STM seems more like the continuous process postulated by signal-detection theory with the individual capable of varying his criterion over a wide range. Of course, probably the data could also be accounted for by various low-threshold models as well (see Norman, 1964); however, the present experiment is too crude to make such a test very feasible.

There are, of course, various objections that could be raised about the present study and its conclusions. First, it could be argued that signal-detection theory has nothing to do with paired associates, or that ROC curves have no place in STM. Admittedly, one cannot specify the waveform or bandwidth of an associative bond, but signal-detection theory provides a model not only for the sensory processes but also for *S*'s decision processes. And, if a threshold is invoked as an explanatory concept, it seems a little unfair to deny the possibility of an empirical test.

A second objection is that the postulated threshold applies to recall, not to recognition. Thus, as used by Underwood and Keppel (1962), the high-threshold concept is applied to studies which test STM by recall. However, by now we have collected sufficient data with paired associates so that we can say with a fair degree of confidence that about 25% of the items from SP 1–4 and about 60% of the items from SP 5 would have been recalled correctly had retention been tested by recall (see Murdock, 1963, page 438, Table 5). Therefore, in terms of the argument of Underwood and Keppel (1962), some of the pairs would have been above and some below their postulated "anticipatory threshold" (page 4, Fig. 2). Recognition rather than recall was used in the present experiment simply because it is generally considered to be a more sensitive test of retention.

A third objection is that the results obtained are an artifact of averaging. That is, in many signal-detection studies one obtains intensive data from a few *S*s and plots ROC curves individually; here the data from all 24 *S*s were pooled to obtain the overall curves of Fig. 1. However, in a study of recognition memory for common English words Egan (1958) found that averaging did not appreciably modify the ROC curves for individual *S*s. And, it might be noted, he too found little evidence for a high-threshold concept. In the present experiment we did try to plot ROC curves for individual *S*s, but the data were too variable to be very revealing.

In conclusion, it would seem that the methods of signal-detection theory may provide a useful approach to the study of problems in STM. This theory, or model, is well developed and leads to a number of testable deductions, and its quantitative nature may be a useful antidote to the vague theorizing that is sometimes characteristic of the verbal-learning field. While it may be too much to think of measuring the strength of memory traces in terms of d' and β, still it should be profitable to apply some of the basic concepts of signal-detection theory to short-term memory.

REFERENCES

BROADBENT, D. E., & GREGORY, M. Division of attention and the decision theory of signal detection. *Proc. Roy. Soc. London, Ser. B*, 1963, **158**, 222–231.

EGAN, J. P. Recognition memory and the operating characteristic. Technical Note AFCRC-TN-58-51, 1958, Indiana University, Hearing and Communication Laboratory.

EGAN, J. P., SCHULMAN, A. I., & GREENBERG, G. Z. Operating characteristics determined by binary decisions and by ratings. *J. Acoust. Soc. Amer.*, 1959, **31**, 768–773.

ESTES, W. K. Learning theory. *Annu. Rev. Psychol.*, 1962, **13**, 107–144.

MURDOCK, B. B., JR. Short-term retention of single paired associates. *J. exp. Psychol.*, 1963, **65**, 433–443.

NORMAN, D. A. Sensory thresholds, response biases, and the neural quantum theory. *J. math. Psychol.*, 1964, **1**, 88–120.

POLLACK, I., & DECKER, L. R. Confidence ratings, message reception, and the receiver operating characteristic. *J. Acoust. Soc. Amer.*, 1958, **30**, 286–292.

SWETS, J. A. Is there a sensory threshold? *Science*, 1961, **134**, 168–177.

SWETS, J. A., TANNER, W. P., & BIRDSALL, T. G. Decision processes in perception. *Psychol. Rev.*, 1961, **68**, 301–340.

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

UNDERWOOD, B. J., & KEPPEL, G. One-trial learning? *J. verbal Learn. verbal Behav.*, 1962, **1**, 1–13.

(Received July 8, 1964)