

Strength Theory of Decision Rules and Latency in Retrieval from Short-Term Memory¹

DONALD A. NORMAN

University of California, San Diego, California 92037

AND

WAYNE A. WICKELGREN

Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

A simple strength theory is studied for retrieval in recognition, multiple-choice and recall tests of verbal short-term memory. The primary assumptions of the theory are: (a) that the same memory system is used, regardless of the manner of testing retention, (b) that in recognition, subjects respond "yes" if the strength of the test item exceeds a criterion (criterion decision rule), and (c) that in multiple-choice or recall, subjects choose the response with the greatest strength (maximum decision rule). The results support these basic retrieval assumptions, but suggest that the decay and noise in retrieval may not always be constant for different methods of testing retrieval. Response latencies were studied as a subsidiary measure of trace strength.

This paper is concerned with three questions about retrieval from verbal short-term memory (STM). First, are the same memory traces used when memory is tested by recognition, multiple-choice and recall tests? Second, what decision rule transforms memory traces into responses for each different way of testing STM? Third, can confidence judgments and response latency supplement choice as measures of trace strength?

In this paper we ignore how material gets into memory (acquisition) and how it deteriorates (decay). Instead, we concentrate on the retrieval (decision) process. The fundamental properties of our strength theory of memory have been presented previously (Wickelgren and Norman, 1966). The decision rules for retrieval which we discuss in the present paper are completely analogous to the decision rules in signal detection theory (Green and Swets, 1966; Swets, Tanner, and Birdsall, 1961). Bower

¹ This work was supported by grants from the National Institutes of Health, U. S. Public Health Service: MH-08083-02 to the Center for Cognitive Studies, Harvard University, and MH-08890-03 to M.I.T. Further aid was received from grant NsG 496 to M.I.T. from the National Aeronautics and Space Administration.

(1967) has also considered the problem of the relationship between different methods of testing retrieval, but his multicomponent theory of memory is at a different level from strength theory and outside the scope of the present paper.

Psychologists differ in their use of the term "STM." For us, STM refers to the memory of the most recently presented material, when the interval between presentation and test is filled with other items that have to be learned or attended to very carefully. STM appears to have a capacity of about 4-5 items or a time constant of 1-10 sec. We take no stand on the issue of whether decay is a function of time or the number of intervening items, although one of us has argued that decay is a function of the number of intervening items (Norman, 1966; Waugh and Norman, 1965).

In the present study we are concerned with STM for any one of four digit-digit paired associates assessed after a delay ranging from 3.5 to 6.5 sec. What the relationship might be between this type of memory and ("intermediate-term") memory for items assessed after filled delays of 15 sec or longer is impossible to say at the present time. Thus, we may well be studying a different memory trace from that studied by Green and Moses (1966) and by Kintsch (1966), who studied the relationship between "yes-no" recognition and multiple-choice tests of memory with delay intervals of 10 min and 20 sec, respectively.

RECOGNITION

In recognition, the subject is asked to decide whether the single test item was presented to him earlier. In this task, strength theory assumes that the subject considers only the strength of the memory trace associated with the test item. In particular, we assume that the subject compares the trace strength of the test item with a criterion, responding "yes" if the strength exceeds the criterion and "no" otherwise. There are two types of test items, "old" (a test item that appeared in the list for that trial) and "new" (one that did not appear in the list). Strength theory assumes that there may be noise in the value of the trace strength, the location of the criterion, or both. We assume that this noise is normally distributed. Thus, trace strength is a random variable with mean t_o (for an "old" test item) or mean t_n (for a "new" test item) and with standard deviation σ_o or σ_n , respectively. The criterion is normally distributed with mean c and standard deviation σ_c .

With these assumptions it is now possible to solve for the probability of responding "yes" to an old or new test item. Plotting the probability of correctly recognizing an old test item against the probability of falsely recognizing a new test item, for different values of c , yields the memory operating characteristic (OC). When plotted on normal-normal probability paper, the OC is a straight line with slope b and horizontal intercept d , where:

$$b = \frac{(\sigma_n^2 + \sigma_c^2)^{1/2}}{(\sigma_o^2 + \sigma_c^2)^{1/2}},$$

and

$$d = \frac{t_o - t_n}{(\sigma_n^2 + \sigma_c^2)^{1/2}}.$$

The derivation of the slope and intercept parameters is well known. It follows directly from decision theory (see Green and Swets, 1966) as applied to recognition memory (see Egan, 1958, and Wickelgren and Norman, 1966).

MULTIPLE-CHOICE

In multiple-choice, the subject is presented with a list of alternative test items and is asked to decide which one was presented to him earlier. In this task, we assume that the subject chooses the test item with the greatest trace strength. It is possible to predict the probability of a correct response in multiple-choice or recall experiments from the parameters of the OC for recognition, provided certain assumptions are satisfied: (a) The means of the strength distributions in multiple-choice are the same as the means for the corresponding strength distributions in recognition. This is a strong assumption since it is quite possible that there is greater degradation of the memory traces during retrieval, when several traces must be examined. (b) The noise affecting the values of the trace strength is the same in the multiple-choice situation as in recognition. This is also a strong assumption since the noise in trace retrieval may increase with the number of traces retrieved. (c) The strength distributions of the items being compared in multiple-choice are uncorrelated. (d) There is a criterion (bias) noise associated with each of the m alternatives in multiple-choice, and this noise is equal to the criterion noise in recognition.

If, in addition, the recognition OCs have unit slope (i.e., $\sigma_n = \sigma_o$) then the probability of a correct response in multiple choice can be determined from the recognition d values using the tables of Elliott (1964).

RECALL

In recall, the subject is asked to generate one of the items presented to him earlier, in the absence of an explicit list of alternative items. When the number of response alternatives, m , is small, and well known to the subject, then we treat the prediction of recall from recognition in exactly the same way as the prediction of m -alternative multiple-choice.

CRITERION DECISION RULE FOR RECOGNITION MEMORY

There are essentially two aspects of the criterion decision rule for recognition. The first aspect is a sort of "independence from irrelevant strengths" assumption. In paired-associate recognition memory, this asserts that what is judged is the strength of

association between the test stimulus and the test response. The strengths of other associations involving the test stimulus and test response are irrelevant to the decision. This strong assumption appears to be valid for recognition memory, at least in some situations (Wickelgren, 1967).

The other aspect of the criterion decision rule is the assumption that the decision rule is deterministic, with probabilistic behavior resulting from the noise in the strengths and/or criteria. There is no direct way to distinguish the predictions of such a decision rule from the predictions of Luce's (1959, 1963) ratio rule, which has a probabilistic decision rule operating on real-variable strengths, but the criterion rule is associated with a simpler law of trace decay (Wickelgren and Norman, 1966).

EQUIVALENCE OF THE MEMORY TRACES

There is actually considerable evidence to support the assumption that the same memory system is employed in both recognition and recall, though there is really no evidence on the relation between the memory system used in recognition and that used in multiple-choice tests of STM.

Qualitative support for the assumption of a common memory system for both recognition and recall is provided by experiments showing that errors in STM tend to be phonetically similar to the correct item for both recall (Conrad, 1964; Wickelgren, 1965a, 1965b, 1966a) and recognition (Wickelgren, 1966b), and that repetition, primacy, and recency affect recall and recognition tests of STM in similar ways. Quantitative support for the assumption of a common memory system for recall and recognition is provided in a study by Norman (1966) which compared the rates of decay for the two types of tests, using strength theory to transform response probabilities into trace strengths in both cases. The rate of decay in verbal STM was quite similar for both recall and recognition.

RESPONSE LATENCY

The basic assumption is that response latency is related to trace strength in exactly the same way that confidence judgments are related to trace strength, namely, that extreme values of latency or confidence reflect extreme values of strength or strength difference.

In recognition, we have assumed that only the strength of the test item determines the response. When the strength of the test item is close to the criterion, we assume that the decision is more difficult, leading to a longer latency. In particular, we assume that latency for "yes" responses is monotonically decreasing with trace strength, and latency for "no" responses is monotonically increasing with trace strength. Thus, low values of strength are associated with short-latency "no" responses, slightly higher values with long-latency "no" responses, still higher values with long-latency "yes" responses, and the highest strength values with short-latency "yes" responses. In

effect, we can treat each separate response latency for "yes" and "no" judgments in exactly the same way that we treat each confidence level for a "yes" and "no" judgment, letting each latency (or confidence level) be considered as a different criterion for choosing a response. The equivalence of confidence ratings and latencies is shown in Fig. 1, where we show possible locations for response criteria in terms of both confidence levels and latencies (the particular values used in Fig. 1 are taken from the recognition results for subject PC in the experiment reported in this paper).

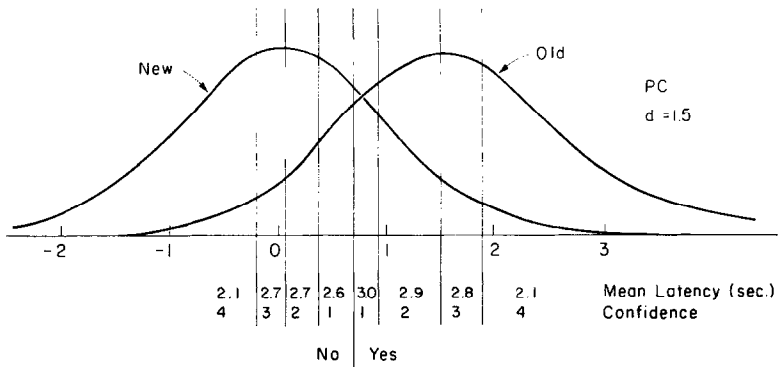


FIG. 1. Strength distributions of old and new items for PC with criteria for confidence levels and mean latencies of "yes" or "no" responses in recognition for each level of confidence.

For a multiple-choice task with two alternatives it is natural to assume that the difficulty of the choice, and hence the latency, is monotonically related to the difference between the two trace strengths being compared. One can then make the assumption that response latency in two alternative multiple-choice has exactly the same relationship to the difference in strength between the two test items as latency in recognition has to the difference between the strength of the (single) test item and the criterion.

For multiple-choice or recall with more than two alternatives, response latency, like confidence, is undoubtedly not determined by the value of a single strength or strength-difference. Both confidence and latency are probably determined by some function of the different strengths that are logically relevant to the choice. However, one would expect confidence and latency to be correlated in the same way as assumed for recognition and two-alternative multiple-choice, that is, short-latency responses, whether correct or incorrect, would be associated with high confidence and long-latency responses with low confidence.

The obvious empirical test of the above assumptions is to determine whether confidence and latency are negatively correlated for each of the four methods of testing retrieval. Tests of whether confidence and latency are measures of trace strength in the manner specified here are provided by the *a posteriori* probability

functions obtained for each. For a definition of the *a posteriori* probability function, see Nachmias and Steinman (1963) or Norman and Wickelgren (1965).

In addition, for recognition and MC-2, it is possible to plot OCs using latencies in combination with the response, in exactly the same manner as one uses confidence judgments in combination with the response. Such latency OCs will have one point in common with the confidence OCs, namely, the "yes-no" point in the case of recognition and the "correct-incorrect" point in the case of two-alternative multiple-choice. Whether the other points of a latency OC correspond to the other points of a confidence OC depends on whether the noise affecting latencies is comparable in magnitude to the noise affecting confidences. If the noise in latencies exceeds the noise in confidences, then the latency OC will lie below the confidence OC at all other points, and vice versa.²

This raises another point. The OC will be a straight line on normal-normal coordinates with a slope which reflects the relative variances of the underlying strength-minus-criterion distributions, only if the variability of the location for each criterion is comparable in magnitude. If OCs determined from response latencies are to be straight lines (on normal-normal coordinates), then each latency which is considered as a criterion) must have equal noise. Moreover, this noise must be equal to the noise in the "yes-no" criterion. The assumption that noise affects all criteria and latencies in the same way would not be necessary if we could assess, independently, the noise levels in the different criteria. Unfortunately, we cannot. However, if the latency or confidence OCs bend down on one or both sides of the "yes-no" point, it will certainly suggest that the variation in the latency or confidence criteria on that side or sides is greater than the variation in the "yes-no" or choice criterion. If the OCs bend up around this point, it suggests the opposite (unlikely) conclusion.

EXPERIMENT

Procedure. Subjects had to remember a list of four digit pairs, preceded and followed by an interference task. Four types of retention tests were presented to the subjects: recognition, recall, and multiple choice with two and four alternatives (MC-2 and MC-4, respectively). The experimental situation was chosen so that all task variables remained constant from one trial to another with the exception of the method of test and the pair tested. Thus, subjects were unaware of the type of test to be presented until the time of the test. The same set of four stimulus digits and the same set of four response digits were employed on each trial. Only the pairing was changed. This procedure minimized item strength (availability), as a factor in the experiment.

The entire experiment was controlled by a digital computer which prepared and presented the stimulus lists, accepted the subjects' responses, scored, and timed them. The stimuli were presented in a series of nine windows (1.5 × 1.75 in.) arranged horizontally, approximately

² This type of latency OC is quite different from that used by Carterette, Friedman and Cosmides (1965) for their latency data. At present, it is not possible to say what relationship, if any, there might be between them.

4 ft from the heads of the subjects. In the discussion that follows, the nine windows of the display will be referred to as W1 through W9, numbering the windows from left to right.

Stimulus List. The stimulus list consisted of four pairs of digits presented one pair at a time for exactly 1 sec in W4 and W5. The left member of the pair was always selected from the digit set 2, 3, 4, and 5; the right member of the pair was always selected from the set 6, 7, 8, and 9. No pair of numbers in the 4-pair set was ever repeated in a trial, so each of the eight possible digits was always presented exactly once in each list of four digit-pairs. Aside from these restrictions, the ordering of the digit pairs varied randomly from trial to trial.

Interference Lists. Immediately before and after presentation of the stimulus list the subjects received an interference task. Single letters of the alphabet were presented for 750 msec each in W4. As the subjects saw each letter, they were to recite aloud the letter in the alphabet that followed the one presented. This task is very difficult at the rate of presentation used. Two letters preceded the stimulus list and four letters followed it. There was no pause between the interference lists and the stimulus list.

Tests. Subjects were able to tell which test was being given by the nature of the test item presentation. The subject was first shown some items from the right-hand part of the digit pairs and then one of the left-hand digits. Their job was to decide which of the right digits had occurred with the left digit. The only possible right-hand digits were the numbers 6, 7, 8, and 9, which were presented in the test in window positions W6, W7, W8, and W9, respectively. Thus, even if the only right test digit being shown was the number 7, it would be shown in its assigned window, W7. The left-hand member of the test was always shown in W4. The right-hand digits were presented without delay following the last interference list. The left-hand digit followed the right-hand digits by 500 msec, but without disrupting the presentation of the right-hand test digits. All digits then stayed on for 5 sec to allow the subjects time to respond. Timing of the response started as soon as the left-hand digit was displayed, and stopped when the subject made his "yes-no" or choice response.

Recognition. One left and one right member of the pair was presented. The subjects were asked to respond by pushing a button marked *y* (yes) if the pair had occurred in the stimulus list, and one marked *n* (no) if it had not.

MC-2. One left and two possible right members were presented, with one left-right pairing having occurred in the list. The subjects were asked to select the right-hand digit that had been paired with the left-hand one by pushing the button marked with that number.

MC-4. The same as MC-2 with the exception that all four possible right-hand digits were shown to the subjects.

Recall. The same as MC-4 except that none of the right-hand digits were presented to the subject. The recall condition of this experiment is logically identical to the MC-4 condition, though it may differ psychologically.

Confidence Judgments. After the subjects had made their decisions they were asked to indicate how confident they were that their response had been correct by pushing a button. When they were very confident, they were asked to push the button marked "4." When they had little confidence (they had guessed), they were asked to push the button marked "1." They were to use the buttons marked "2" and "3" for intermediate levels of confidence.

Design. In recognition tests, the test pair of digits was composed of one of the four left digits and one of the four right digits from the pairs in the presentation. Thus, there were 16 possible

test pairs, 4 which had been previously presented to the subjects (correct pairs) and 12 which were novel pairings for that trial (incorrect pairs).

In MC-2, the left digit of the test pair could come from any one of the four left digits of the list presentation, but of the two right digits, one had to make a correct pair, and was therefore constrained by the choice of the left-hand digit, while the other right-hand digit could come from any of the other three possibilities. Thus, there were 12 possible conditions for MC-2 tests.

In MC-4 and recall, there were only four possible test conditions, representing a test of any of the four pairs presented in the trial.

Each of the 36 conditions was presented once to each subject in each block of trials. The order of the tests varied randomly from block to block. Each block was preceded by five practice trials, although the subjects were not aware of this. Subjects were required to take a 5-min break between blocks, and always did an intergral number of blocks on any 1 day, usually three, and sometimes as many as five. It took approximately 20 min to finish a block. Two subjects had one day of (unscored) practice, and the other two subjects had 2 days of practice.

Subjects. Subjects were four male undergraduates at Harvard University, selected only for their availability during the 60 days of the experiment. They were paid \$1.50/hour for their services.

TABLE 1
PROBABILITIES OF "YES" RESPONSES IN RECOGNITION

Condition ^a (<i>i, j</i>)	Subject			
	ES	JS	PC	SS
1,1	.91	.77	.84	.22
1,2	.10	.11	.32	.17
1,3	.04	.17	.24	.10
1,4	.05	.15	.22	.06
2,1	.07	.06	.26	.16
2,2	.91	.71	.70	.21
2,3	.08	.05	.24	.15
2,4	.03	.17	.22	.07
3,1	.06	.11	.26	.11
3,2	.09	.12	.26	.13
3,3	.88	.75	.76	.20
3,4	.02	.15	.24	.11
4,1	.02	.06	.22	.14
4,2	.07	.14	.20	.16
4,3	.05	.18	.16	.14
4,4	.91	.75	.82	.31
<i>N</i> ^b	100	65	50	99

^a Condition (1, *j*) means the left member of the test pair was the same as the left member of the *i*th pair of the preceding 4-pair list and the right member of the test pair was the same as the right member of the *j*th pair of the preceding 4-pair list.

^b *N* is the number of trials on which each probability was estimated.

RESULTS AND DISCUSSION

RESPONSE PROBABILITIES

The probabilities of "yes" responses in each recognition condition and the probabilities of correct responses in MC-2, MC-4, and recall conditions are presented in Tables 1 and 2. The level of performance varies greatly over the four subjects, but for a given subject in a given retrieval condition, performance is not very different for the different serial positions of the test stimulus. Furthermore, the false recognition rate for incorrect test pairs is not very different for different serial positions of the incorrect test response in relation to the test stimulus, nor are the differences that do exist very consistent over the different subjects. The probabilities of correct choice in MC-2 are not very different for different serial positions of the incorrect alternative response. For these reasons, many of the subsequent analyses will be on data for a given subject lumped into five categories: Correct recognition, false recognition, MC-2, MC-4, and recall.

TABLE 2
PROBABILITIES OF CORRECT RESPONSES IN MC-2, MC-4, AND RECALL

MC-2 condition (<i>i, i</i>) vs (<i>i, j</i>)		Subject				MC-4 condition	Subject			
		ES	JS	PC	SS		ES	JS	PC	SS
1,1	1,2	.97	.86	.80	.54	1	.82	.58	.62	.28
1,1	1,3	.97	.88	.80	.64	2	.82	.74	.52	.24
1,1	1,4	.96	.86	.76	.53	3	.90	.80	.56	.35
2,2	2,1	.91	.89	.82	.55	4	.89	.82	.54	.32
2,2	2,3	.96	.85	.82	.61	<i>N</i>	100	65	50	99
2,2	2,4	.97	.92	.80	.56	Recall condition				
3,3	3,1	.95	.78	.88	.60	1	.91	.80	.60	.36
3,3	3,2	.94	.92	.78	.58	2	.84	.57	.46	.30
3,3	3,4	.92	.86	.68	.51	3	.87	.55	.60	.32
4,4	4,1	.97	.89	.80	.58	4	.86	.82	.64	.45
4,4	4,2	.96	.94	.68	.58	<i>N</i>	100	65	50	99
4,4	4,3	.96	.91	.76	.58					
<i>N</i>		100	65	50	99					

LATENCY AND CONFIDENCE *a Posteriori* PROBABILITIES

The *a posteriori* probability function for recognition is the probability that, when any specific response (Y_4, \dots, N_4) is chosen by a subject, the test item was, in fact, old. If subjects are ordering the criteria for the different levels of confidence along the strength dimension, as assumed by the theory, then the empirical *a posteriori* prob-

abilities should decrease from Y4 to Y1 and from N1 to N4. Figure 2 presents the empirical *a posteriori* probabilities for recognition and demonstrates that, with a few minor exceptions, the criteria for the confidence responses are ordered in the expected manner on the strength dimension. Latency *a posteriori* probabilities exhibit a similar function. Thus, both confidence and latency reflect trace strength in recognition.

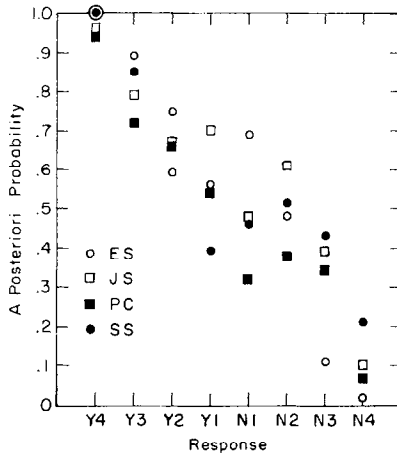


FIG. 2. Confidence *a posteriori* probabilities for recognition.

For MC-2, MC-4, and recall, we computed the empirical *a posteriori* probabilities that, when a particular level of confidence was chosen by a subject, the choice the subject made was correct. Again, *a posteriori* probability was a monotonic increasing function of level of confidence and a monotonic decreasing function of latency.

Table 3 presents the mean latency with which each of the 4 subjects made a "yes-no" response in the case of recognition or a "correct-incorrect" choice in the case of MC-2, MC-4, and recall. The mean latencies are given separately for each level of confidence. It is clear from Table 3 that latency and confidence are highly negatively correlated, that is, short latency responses go with high levels of confidence.

LATENCY AND CONFIDENCE OPERATING CHARACTERISTICS

Another way to talk about the information concerning trace strength that is transmitted by latencies judgments is to ask if the criterion placement for a given latency on the strength scale has greater noise in it than the criterion placement for a given confidence level. Information concerning the relative noise in "yes-no," confidence, and latency criteria can be obtained by examination of the empirical OCs for recognition using either confidence or latency in conjunction with the "yes-no" decision.

TABLE 3
 LATENCY (sec) FOR RESPONSES IN RECOGNITION, MC-2, MC-4,
 AND RECALL AS A FUNCTION OF LEVEL OF CONFIDENCE

Subject	Recognition: old test item							
	Y4	Y3	Y2	Y1	N1	N2	N3	N4
ES	1.48	1.95	2.29	2.24	2.72	2.25	2.06	1.41
JS	1.70	2.52	2.48	2.92	2.49	2.60	2.56	1.76
PC	2.11	2.94	2.91	2.94	2.48	2.80	2.72	2.20
SS	1.43	1.67	2.38	2.47	1.91	1.57	1.35	1.15
Subject	Recognition: new test item							
	Y4	Y3	Y2	Y1	N1	N2	N3	N4
ES	2.13	2.12	2.31	2.68	2.08	2.32	1.88	1.30
JS	2.50	2.43	2.64	3.53	2.38	2.35	2.11	1.48
PC	2.07	2.60	2.88	3.05	2.73	2.63	2.64	2.07
SS	—	1.57	2.18	2.44	1.96	1.57	1.38	1.62
Subject	MC-2							
	C4	C3	C2	C1	I1	I2	I3	I4
ES	1.43	1.95	2.57	2.80	2.84	2.55	2.13	1.73
JS	1.72	2.36	2.64	2.51	2.97	2.68	2.65	2.15
PC	2.27	2.77	3.04	3.07	3.10	3.07	2.88	2.30
SS	1.55	1.71	1.78	2.12	2.15	1.71	1.60	1.23
Subject	MC-4							
	Y4	Y3	Y2	Y1	N1	N2	N3	N4
ES	1.53	2.08	2.74	3.33	2.99	2.80	1.82	1.23
JS	1.74	2.64	2.89	3.03	3.03	3.02	2.92	1.91
PC	2.31	3.33	3.27	3.15	3.40	3.59	2.64	2.79
SS	1.80	2.27	1.82	2.37	2.32	1.90	1.76	—
Subject	Recall							
	Y4	Y3	Y2	Y1	N1	N2	N3	N4
ES	1.14	1.52	2.17	2.49	2.57	2.54	1.82	1.03
JS	1.31	2.02	2.30	2.56	2.49	2.24	1.92	1.24
PC	1.81	2.21	2.41	2.73	2.58	2.71	2.56	2.04
SS	1.06	1.31	1.32	1.61	1.97	1.35	1.12	2.00

Similarly, we can examine the OCs for MC-2 using either confidence or latency in conjunction with the "correct-incorrect" choice of response.

Figure 3 presents both confidence and latency recognition OCs (on a normal-normal plot) for each subject separately, lumping all the old test item conditions together to determine vertical-axis values, and lumping all the new test item conditions together for the horizontal axis. Figure 4 presents both confidence and latency OCs for each subject, lumping all 12 MC-2 conditions together. It is clear from these

figures that the latency OCs are systematically below the confidence OCs for both recognition and MC-2 at all points except the “yes–no” or “correct–incorrect” point, at which the OCs must coincide. The same conclusion is reached by examination of the OCs obtained without lumping the various conditions together, verifying that the peculiar shapes of the latency OCs are not the result of averaging.

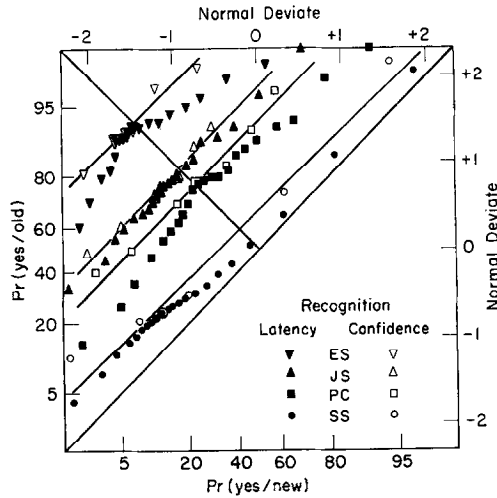


FIG. 3. Confidence and latency OCs for recognition.

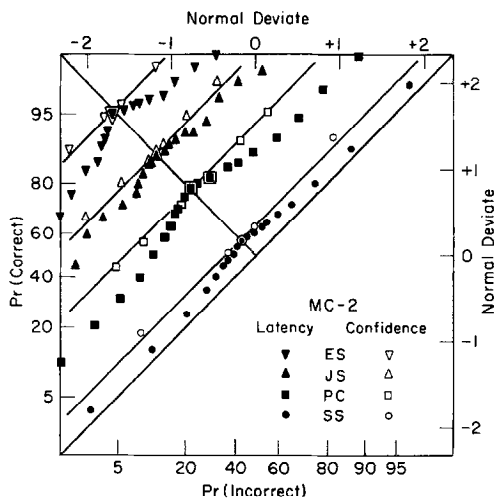


FIG. 4. Confidence and latency OCs for MC-2.

Several points should be noted about the OCs in Figs. 3 and 4. First, only one confidence OC (SS-recognition) is at all deviant from a straight line on a normal-normal probability plot. The confidence OCs for the other three subjects are quite consistent with the assumption of normally distributed noise.

Second, the "yes-no" and "correct-incorrect" points are certainly not systematically above the best-fitting straight line that could be drawn through the "confidence" points. This indicates that there is no more noise in the placement of a confidence criterion than in the placement of a "yes-no" or response-choice criterion in recognition or MC-2, respectively.

Third, the slopes of the best-fitting lines for the confidence OCs in Fig. 3 are very close to unity, indicating that the noise in the false recognition conditions was virtually identical to that in the correct recognition conditions.

Fourth, the present manner of generating the MC-2 OCs in Fig. 4 guarantees symmetry about the negative diagonal, so this aspect of Fig. 4 should be ignored. Only the slope of an OC on one side of the negative diagonal is of interest. These are, however, quite close to unity for the confidence OCs in MC-2, in accord with the prediction of strength theory.

RELATION BETWEEN RECOGNITION, MULTIPLE-CHOICE, AND RECALL PROBABILITIES

Throughout this section, recall and MC-4 will be averaged together and referred to as MC-4. The justification for this is that there were no systematic differences between them in the probability of correct choice (see Table 2). There were large unsystematic differences between them, which were not consistent from subject to subject. However, even these differences in a single subject were never consistently in favor of one or the other over all four serial positions. There is some indication of a greater primacy effect in recall than in MC-4, but is not a clear-cut difference. Thus, to reduce the noise in the data we average the probability of correct choice for recall and MC-4, but we do this separately for each serial position ($i = 1, 2, 3, 4$). Whatever the difference between the decision processes in recall and MC-4, these differences did not result in a clear-cut difference in accuracy in the present experiment.

To reduce noise in MC-2 and because the distractors from various serial positions did not seem to have a large, consistent effect on the probability of correct choice (see Table 2), we have averaged together the data for all the correct and incorrect conditions for a given serial position of the test stimulus (i).

For the same reasons we have averaged together the data for all the "new" item recognition conditions for a given serial position of the test stimulus (i).

Having done this, we now use the recognition data to make predictions of the probabilities of correct choice in MC-2 and MC-4 (including recall). We use the d value for recognition to compute the expected probability of correct choice in MC-2 and MC-4, using the tables of Elliott (1964, p. 682). In Figs. 5A and B, observed

probabilities of correct choice in MC-2 and MC-4 (p_2 and p_4) are plotted against the expected probabilities (ep_2 and ep_4). The expected probability of correct choice in MC-4 can also be derived from the data for MC-2 using the tables of Elliott, and a plot of p_4 against this expected probability, ep'_4 , is shown in Fig. 5C.

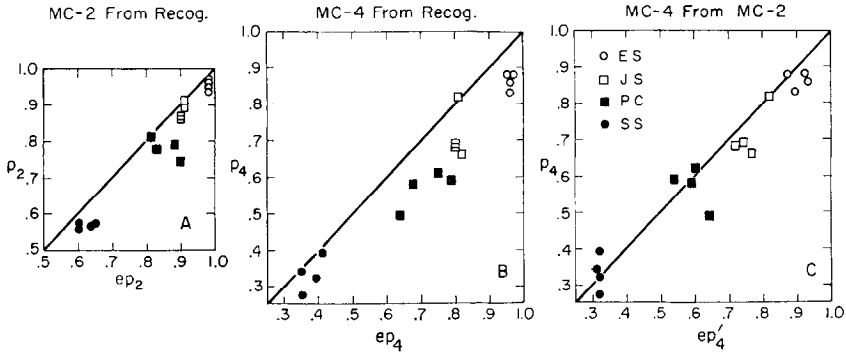


FIG. 5. Plots of obtained (p_2, p_4) vs expected (ep_2, ep_4, ep'_4) probabilities in MC-2 and MC-4, using the strength theory decision rules: (left), MC-2 is predicted from recognition; (center) MC-4 is predicted from recognition; and (right), MC-4 is predicted from MC-2.

There is not much variation in the expected and observed probabilities for a given subject over the different serial positions, but there is considerable variation in the level of performance shown by the four different subjects. Thus, we have some idea of the adequacy of simple strength theory for predicting multiple-choice and recall from recognition and for predicting MC-4 from MC-2 over a broad range of probabilities of correct performance.

Although there is some variability in the relationship between expected and observed probabilities in Fig. 5, the conclusion is clear. Simple strength theory overpredicts the probability of correct performance in MC-2 and MC-4 on the basis of the recognition data. The discrepancy is not too great for MC-2, but somewhat larger for MC-4. The prediction of MC-4 from MC-2 is fairly good, although there is a slight overprediction for the best two subjects.

What do we conclude from these results? In view of the previous evidence in favor of a common memory system for recognition and recall, and the present evidence of an extremely strong correlation between recognition, multiple-choice, and recall performance over different individuals, it is unlikely that the fault lies in the assumption of a common memory. Previous evidence strongly supports the criterion decision rule for recognition memory, and both previous studies and the present study produced OCs that were quite consistent with the normality assumptions for recognition and MC-2. Furthermore, distribution-free methods of predicting multiple-choice

from recognition (see Green and Swets, 1966, pp. 45-49) yielded plots essentially identical to Figs. 5A and B.

Given that one is satisfied with the criterion rule for recognition memory, the maximum rule for multiple-choice and recall has a certain intrinsic plausibility that makes it seem very unlikely to be at fault, unless the basic theoretical concepts of strength theory are incorrect. Furthermore, the relatively accurate prediction from MC-2 to MC-4 argues that the maximum rule is correct for multiple-choice and recall. Thus, the overprediction of multiple-choice and recall from recognition probably results from one or more of the following: (a) negative correlation of the distributions, (b) greater decay during retrieval with increasing number of traces retrieved, or (c) greater noise during retrieval (in strengths or criteria) with increasing number of traces retrieved. There is no way to decide among these three possibilities on the basis of the present data.

LATENCIES FOR DIFFERENT RETRIEVAL CONDITIONS

Table 4 presents the means and standard deviations of the latency distributions for each binary response for each retrieval condition in the present experiment.

TABLE 4
MEANS AND STANDARD DEVIATIONS OF THE LATENCY DISTRIBUTIONS
FOR THE BINARY RESPONSES IN EACH RETRIEVAL CONDITION

Condition	Response	Mean (sec)				SD (sec)			
		ES	JS	PC	SS	ES	JS	PC	SS
Correct Recog.	Yes	1.64	2.00	2.50	2.01	.59	.77	.82	.81
	No	2.17	2.31	2.60	1.52	.92	.78	.71	.68
False Recog.	Yes	2.29	2.64	2.79	2.22	.78	.82	.86	.85
	No	1.44	1.63	2.34	1.53	.58	.65	.79	.70
MC-2	Corr.	1.61	1.89	2.57	1.78	.63	.69	.77	.70
	Incorr.	2.42	2.62	2.87	1.75	.84	.72	.83	.70
MC-4	Corr.	1.89	2.19	2.80	1.98	.84	.85	.93	.83
	Incorr.	2.65	2.89	3.31	2.00	.82	.77	.75	.84
Recall	Corr.	1.39	1.70	2.13	1.35	.73	.82	.83	.67
	Incorr.	2.32	2.14	2.60	1.45	1.05	.90	.87	.79

Note that response latency increases from recognition to MC-2 to MC-4, but the entire increase amounts to about 0.1 sec/additional alternative. That is to say, correct or incorrect responses in MC-2 have a latency about 0.1 sec longer than correct or

incorrect responses, respectively, in recognition, while correct or incorrect responses in MC-4 have a latency about 0.2 sec longer than correct or incorrect responses, respectively, in MC-2. This is not a very large increase in response latency. Thus, while there may be somewhat greater decay of the memory traces in retrieval for MC-4 than for MC-2 and, in turn, than for recognition, the effect is not likely to be too large. However, this latency difference is qualitatively consistent with the explanation that the overprediction of performance shown in Fig. 5 may result from decay during the retrieval period.

According to the theory, recall requires four traces to be retrieved and a response set of four. Recognition has only one trace to be retrieved, and only two responses to choose from. Yet latency in recall is less than that in recognition. Moreover, the latencies for recall are shorter than for MC-4, even though the two conditions are equivalent on logical grounds. This result suggests that rather different retrieval methods are being employed in recall than in multiple-choice. It does not mean that the maximum rule cannot hold for both recall and multiple-choice. After all, the maximum rule is at a fairly high level of abstraction, saying nothing about the mechanism by which the maximum associate is chosen. The fact that in this experiment, the probability of correct response in recall was not systematically different from that for MC-4, suggests that, whatever the differences in mechanism, the two conditions may be described by the same assumptions at the abstract level of strength theory.

REFERENCES

- BOWER, G. A multicomponent theory of the memory trace. In K. W. Spence and J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory*. Vol. I. New York: Academic Press, 1967.
- CARTERETTE, E. C., FRIEDMAN, M. P., AND COSMIDES, R. Reaction-time distributions in the detection of weak signals in noise. *The Journal of the Acoustical Society of America*, 1965, **38**, 531-542.
- CONRAD, R. Acoustic confusions in immediate memory. *British Journal of Psychology*, 1964, **55**, 75-84.
- EGAN, J. P. Recognition memory and the operating characteristic. *Indiana Univer., Hearing and Communication Lab.*, AFCRC-TN-58-51, AD-152650, 1958.
- ELLIOTT, P. B. Tables of d' . In J. A. Swets (Ed.), *Signal detection and recognition by human observers*. New York: Wiley, 1964.
- GREEN, D. M., AND MOSES, F. L. On the equivalence of two recognition measures of short-term memory. *Psychological Bulletin*, 1966, **66**, 228-234.
- GREEN, D. M., AND SWETS, J. A. *Signal detection and psychophysics*. New York: Wiley, 1966.
- KINTSCH, W. An experimental analysis of single stimulus tests and multiple choice tests of recognition memory. Paper presented at the Psychonomic Society Convention, 1966.
- LUCE, R. D. *Individual choice behavior*. New York: Wiley, 1959.
- LUCE, R. D. Detection and recognition. In R. D. Luce, R. R. Bush, and E. Galanter (Eds.), *Handbook of mathematical psychology*. Vol. I. New York: Wiley, 1963. Pp. 103-189.

- NACHMIAS, J., AND STEINMAN, R. M. Study of absolute visual detection by the rating-scale method. *Journal of the Optical Society of America*, 1963, **53**, 1206-1213.
- NORMAN, D. A. Acquisition and retention in short-term memory. *Journal of Experimental Psychology*, 1966, **72**, 369-381.
- NORMAN, D. A., AND WICKELGREN, W. A. Short-term recognition memory for single digits and pairs of digits. *Journal of Experimental Psychology*, 1965, **70**, 479-489.
- SWETS, J. A., TANNER, W. P., JR., AND BIRDSALL, T. G. Decision process in perception. *Psychological Review*, 1961, **68**, 301-340.
- WAUGH, N. C., AND NORMAN, D. A. Primary memory. *Psychological Review*, 1965, **72**, 89-104.
- WICKELGREN, W. A. Acoustic similarity and intrusion errors in short-term memory. *Journal of Experimental Psychology*, 1965, **70**, 102-108. (a)
- WICKELGREN, W. A. Distinctive features and errors in short-term memory for English vowels. *Journal of the Acoustical Society of America*, 1965, **38**, 583-588. (b)
- WICKELGREN, W. A. Distinctive features and errors in short-term memory for English consonants. *Journal of the Acoustical Society of America*, 1966, **39**, 388-398. (a)
- WICKELGREN, W. A. Short-term recognition memory for single letters and phonemic similarity of retroactive interference. *Quarterly Journal of Experimental Psychology*, 1966, **18**, 55-62. (b)
- WICKELGREN, W. A. Exponential decay and independence from irrelevant associations in short-term recognition memory for serial order. *Journal of Experimental Psychology*, 1967, **73**, 165-171.
- WICKELGREN, W. A., AND NORMAN, D. A. Strength models and serial position in short-term recognition memory. *Journal of Mathematical Psychology*, 1966, **3**, 316-347.

RECEIVED: August 21, 1967