The Spacing Effect: Consolidation or Differential Encoding?[1]

ROBERT A. BJORK AND TED W. ALLEN

University of Michigan, Ann Arbor, Michigan 48104

Theoretical explanations of the spacing effect fall into two classes: those that attribute the advantage of two spaced presentations over two massed presentations to better consolidation of the first presentation, and those that attribute the advantage to better encoding of the second presentation. This paper reports an experimental test of the two classes of theory. Rather than manipulate spacing, the experiment varied the information processing difficulty of the activity interpolated between two presentations of an item. Consolidation-type theories imply decreasing consolidation with increasing difficulty of the interpolated activity. In fact, recall performance following two presentations separated by a difficult task was found to be slightly but consistently better than performance following two presentations separated by an easy task. The outcome thus favors encoding-type theories.

There is evidence from a variety of paradigms that the spacing of repetitions of an item has large and relatively clear effects on both latency and frequency measures of performance (for a review, see Bjork, 1970). Of these effects, the largest and most general is that recall tends to improve with the spacing of repetitions of an item.

A profusion of different models has been proposed to explain the improvement in recall performance that results from increased spacing of presentations. These explanations fall into two classes. One class of explanations attributes the spacing effect to an assumed consolidation process of some kind that occurs during the interval between presentations. The longer the interval between presentations, the greater the consolidation of the

long-term memory storage of the item during the interval. The second class of explanations attributes the spacing effect to an assumed variation in the encoding of a presentation as a function of the time since the preceding presentation. The longer the interval between two presentations, the greater the likelihood that the encoding of the second presentation will be a new encoding, that is, will differ from the encoding of the first presentation.

Both the consolidation and the differential encoding explanations are represented by a number of specific models. The rehearsal-buffer theory proposed by Atkinson and Shiffrin (1968), the trace-consolidation theory proposed by Landauer (1969), and the multi-trace strength theory proposed by Wickelgren (1970) are all examples of the consolidation class. The stimulus fluctuation versions of stimulus sampling theory (Estes, 1955; Izawa, 1967), the multistate Markov models of Bjork (1966), Greeno (1967), Rumelhart (1967) and others, and the encoding variability ideas of Martin (1968) and Melton (1969) are all examples of the differential encoding class.

Certainly the two types of explanation of the spacing effect seem very different: they postulate different processes and they assume a different locus of the spacing effect. The

consolidation idea asserts that the first of two spaced presentations is more effective than the first of two massed presentations, and the encoding variability idea asserts that the second of two spaced presentations is more effective than the second of two massed presentations. In spite of how different the ideas might seem, however, it is very difficult if not impossible to decide between them on the basis of spacing experiments. Rumelhart (1967) has shown, for example, that there are specific versions of the rehearsal-buffer, stimulus fluctuation, and multistate Markov theories that are formally equivalent in their predictions of performance in continuous paired-associate spacing experiments.

This paper reports a new attack on the problem. Rather than investigate the effects of the temporal spacing of two presentations, the present experiment manipulates the difficulty level of the information processing required of the subject during the interval that separates the presentations of the material to be remembered. Consider the two conditions diagrammed below.

P_1 . . . easy . . . P_2 . . . medium . . . Test
P_1 . . . difficult . . . P_2 . . . medium . . . Test

In both conditions the time from the first (P_1) to the second (P_2) presentation, the length of the P_2–Test retention interval, and the difficulty level of the information processing during the retention interval are the same. Only the processing difficulty during the spacing interval is varied.

Consolidation theory predicts better performance on the test in the top condition than on the test in the bottom condition. The easy intervening task should permit more consolidation than should the difficult intervening task. For example, in the Atkinson and Shiffrin (1968) model, the average time the to-be-remembered item remains in the rehearsal buffer during the easy intervening task should exceed the average time an item is resident in the buffer during the difficult task. Hence, the amount of information trans-ferred to long-term storage during the easy task should exceed that transferred during the difficult task. One might, of course, postulate an unusual consolidation process of some kind that depends only on time and is independent of the activity level during the consolidation period. Such a process would predict equal performance in the two conditions.

The predictions of the differential encoding theory are less definite. Performance could be better in the difficult-processing condition in spite of the rehearsal or fatigue-reduction advantages of the easy-processing condition. If increasing the activity level produces an increase in the effectiveness of the second presentation in a fashion analogous to that assumed to occur when the interval is lengthened, performance on the test could be better in the difficult processing condition. For example, from the standpoint of stimulus fluctuation theory, the difficult task might produce a higher rate of fluctuation and, hence, more new elements might be available for sampling at the time of the second presentation. In the multistate Markov models, an item should be less likely in the difficult-processing condition to remain in the short-term state from P_1 to P_2. Under the assumption that the probability of transition to the long-term state is higher from the forgotten state than from the short-term memory state (an assumption for which there is considerable support), the probability of transition to the long-term state at P_2 should be higher in the difficult-processing condition than in the easy-processing condition.

METHOD

The experiment employed a variation of the Brown-Peterson paradigm. Subjects were asked to remember trigrams consisting of three common four-letter nouns. After a first presentation of a trigram, Ss shadowed digits until they were either cued to recall the trigram or were given a second presentation of the trigram, in which case, they again shadowed digits until they were cued to recall.

Subjects. The Ss were 30 undergraduate women at

the University of Michigan. They were paid for their participation.

Materials and apparatus. An experimental session consisted of 60 trials. The same 60 trigrams were presented to each *S*, although the assignment of the trigrams to the various trial types varied for each *S*. The trigrams were constructed from common four-letter nouns under the following constraints: (*a*) no two words in a trigram began with the same letter, (*b*) no two words in a trigram rhymed, (*c*) no word was used in more than one trigram, and (*d*) there were no obvious associations between the words constituting any one trigram.

The apparatus was a high speed (change time less than .05 sec) memory drum. The word trigrams, the digits to be shadowed, and the cues to recall, rest, or get ready for the next trial all appeared in the same window.

Design. Trigrams were given one or two presentations on a trial. On the single-presentation trials, the retention interval was either short or long and the shadowing task during the interval was either easy or difficult. Thus, there were four single-presentation conditions. On the double-presentation trials, the interval between the presentations was either short or long, the shadowing task during the spacing interval was either easy or difficult, and the final retention interval was short or long and was filled with a shadowing task of medium difficulty. Thus, there were eight double-presentation conditions.

Every block of 12 trials during an experimental session contained all 12 conditions in random order. Every *S*, therefore, received five trials on each condition during the 60 trials in a session. Across *S*s, the assignment of particular trigrams to the 12 conditions was counterbalanced to insure that the same trigrams were tested in every condition.

Procedure. Every trial began as follows: the word READY was shown for 2 sec, the drum advanced to a blank space for 1 sec, and a trigram was shown for 2.5 sec. Following the presentation of the trigram, a series of three-digit (easy task) or five-digit (difficult task) numbers were shown at a 1.5-sec rate. The three-digit numbers could be shadowed quite comfortably at a 1.5-sec rate. It was very difficult to shadow the five-digit numbers; *S*s were often unable to shadow the fifth digit in a number before the drum advanced to the next five-digit number. The first shadowing interval was 3 or 12 numbers long (4.5 or 18 sec long in elapsed time). After the first shadowing interval, *S*s were either cued to recall the trigram—the word RECALL was shown for 5 sec—or they were given another 2.5-sec presentation of the trigram. In the latter case, there was a second shadowing interval of 8 or 20 four-digit numbers before the *S* was cued to recall. Following the RECALL instruction in both single- and double-

21

presentation trials, a REST instruction was shown for 10 sec before the next trial started.

Before the experimental session began, each *S* was read a set of instructions and was shown four practice trials. Subjects recalled vocally and their recall was recorded by the experimenter.

RESULTS

The recall of a trigram was scored both in terms of whether the entire trigram was recalled correctly or not, and in terms of whether each individual word in the trigram was recalled correctly or not. The order in which the words in a trigram were recalled was either counted in the scoring (strict scoring) or disregarded (lenient scoring). The recall percentages reported in Table 1 are based on the lenient scoring. The difference between the strict and lenient scoring averaged only about 2% and was very consistent across conditions.

TABLE 1

PERCENTAGES OF INDIVIDUAL WORDS RECALLED INCORRECTLY ON SINGLE-PRESENTATION TRIALS[a]

Shadowing difficulty	Retention interval	
	$P_1 - T = 3$	$P_1 - T = 12$
Easy	24 (46)	34 (55)
Difficult	38 (61)	49 (68)

[a] The percentages of trigrams recalled imperfectly are shown in parentheses.

Performance on single-presentation trials. In Table 1 are shown the error percentages observed in the single-presentation conditions. The single-presentation conditions were included in the design for two reasons: they provide a check as to whether the difficult task was in fact, as intended, more disruptive of retention, and they give a general picture of the strength of the memory trace at the time of the second presentation in the various double-presentation conditions. Table 1 leaves no doubt that the difficult task was more

disruptive of retention than was the easy task. In fact, performance following a short interval of difficult shadowing was slightly worse than performance following a long interval of easy shadowing. There appeared to be no interaction between difficulty level and interval length: an increase in shadowing difficulty produced a 14–15% error increase in the recall of individual words, and an increase in the interval length produced a 9–10% increase in errors. The pattern of effects on the recall of entire trigrams is very similar.

Performance on double-presentation trials. The error percentages observed in the double-presentation conditions are shown in Table 2. The comparisons of primary interest in Table 2

TABLE 2

PERCENTAGES OF INDIVIDUAL WORDS RECALLED
INCORRECTLY ON DOUBLE-PRESENTATION TRIALS[a]

$P_1 - P_2$ Interval	Retention interval	
	$P_2 - T = 8$	$P_2 - T = 20$
Short		
Easy	13 (27)	21 (39)
Difficult	11 (25)	21 (37)
Long		
Easy	8 (18)	18 (30)
Difficult	6 (16)	11 (23)

[a] The percentages of trigrams recalled imperfectly are shown in parentheses.

are between double-presentation conditions that differed only in the difficulty level of the shadowing task during the spacing interval. Although the error percentages were quite close in absolute terms, performance in the difficult-processing conditions appeared slightly but consistently better than performance in the easy-processing conditions. This result is particularly striking in view of the differences in retention levels between the easy and difficult conditions at the time of the second presentation, as revealed by the single-

presentation conditions. The average error percentage in the recall of words at the time of the second presentation was 44% in the difficult conditions and 29% in the easy conditions (from Table 1). When there was a second presentation, however, the average error percentages in word recall were 12 and 15% for the difficult and easy conditions, respectively. From the standpoint of commonsense notions of the memory process, Table 2 makes little sense, given the results in Table 1.

The beneficial effects of spacing presentations are quite apparent in Table 2. Averaged over difficulty level during the spacing interval and length of the retention interval, there were 60% more errors (16.5 vs. 10.3) in word recall when the P_1-P_2 interval was short than there were when the P_1-P_2 interval was long.

DISCUSSION

The results of the present experiment suggest a kind of tradeoff between the length of a distractor task interpolated between two presentations of an item and the difficulty level of the distractor task. A short interval filled with a relatively difficult task simulates the effects of a long interval filled with a relatively easy task, both in terms of performance measured at the end of the interval and, more importantly, in terms of the effects of a repetition at the end of the interval on subsequent performance.

The general theoretical implications of a tradeoff between interpolated task difficulty and task duration seem quite clear: both manipulations influence the effectiveness of a repetition following the interpolated task because they influence the *S*'s encoding of the repetition. Consolidation may be an important process in human short-term verbal memory, but no unelaborated current model of the consolidation type is consistent with the results of the present experiment. A long interval of easy shadowing should lead to more consolidation than a shorter interval of more difficult

shadowing, both because it is longer and because the activity level is lower. It is true that performance in the easy-task conditions in Table 2 is only slightly worse than performance in the difficult-task conditions, but it is certainly not better.

It is one thing to feel quite sure that the spacing effect is primarily a matter of differential encoding rather than trace consolidation, but it is another thing to choose among alternative theories of the differential encoding type. Spacing-difficulty experiments more detailed than the present experiment might provide the basis for choosing among alternative specific models.

As a concluding comment, it seems worth pointing out that one general conceptualization of the learning-memory process, explicit in many theories a few years ago and often still implicit in our thinking now, is remarkably wrong. The general tenets of this conceptualization are (a) memory traces can be thought to vary in strength along some continuum, (b) the strength of a memory trace determines the probability of recall in a direct way, (c) during interference the strength of a trace decreases in an orderly fashion, and (d) a repetition increments the strength of a memory trace according to some inverse function of the current strength of the trace such that the stronger of two traces before repetition is still stronger after repetition.

Such a view of memory predicts that the ordering of the recall probabilities in the single-presentation conditions should determine the ordering of the recall probabilities in the double-presentation conditions. That is, the stronger the memory trace at P_2, the stronger it will be after P_2 and, hence, the higher the final probability of recall. This prediction is not only unsupported in the present experiment, it is almost perfectly wrong. The short-easy and long-difficult conditions provide a dramatic example. Error frequency after a short period of easy shadowing is only one-half of what it is after a long period of difficult shadowing (Table 1). When a second

21*

presentation is given, however, error frequency after a long final retention interval is twice as high in the short-easy condition as in the long-difficult condition (Table 2). Overall, performance in the single-presentation conditions correlates −.85 with performance averaged over the two final retention intervals in the corresponding double-presentation conditions.

It is important to continually remind ourselves that recall probability at any point in time is an imperfect indicator of both the "strength" of the representation of an item in memory and the expected changes in recall probability following a repetition. An adequate theory of memory may depend on a multifaceted representation of the memory trace quite unlike anything assumed by current theories.

REFERENCES

ATKINSON, R. C., & SHIFFRIN, R. M. Human memory: A proposed system and its control processes. In K. W. Spence and J. T. Spence (Eds.), The psychology of learning and motivation: Advances in research and theory. Vol. II. New York: Academic Press, 1968.

BJORK, R. A. Learning and short-term retention of paired-associates in relation to specific sequences of interpresentation intervals. Technical Report No. 106, 1966, Institute for Mathematical Studies in the Social Sciences, Stanford University.

BJORK, R. A. Repetition and rehearsal mechanisms in models of short-term memory. In D. A. Norman (Ed.), Models of memory. New York: Academic Press, 1970.

ESTES, W. K. Statistical theory of spontaneous recovery and regression. Psychological Review, 1955, 62, 145–154.

GREENO, J. G. Paired-associate learning with short-term retention: Mathematical analysis and data regarding identification of parameters. Journal of Mathematical Psychology, 1967, 4, 430–472.

IZAWA, C. Function of test trials in paired-associate learning. Journal of Experimental Psychology, 1967, 75, 194–209.

LANDAUER, T. K. Reinforcement as consolidation. Psychological Review, 1969, 76, 82–96.

MARTIN, E. Stimulus meaningfulness and paired associate transfer: an encoding variability hypothesis. Psychological Review, 1968, 75, 421–441.

MELTON, A. W. Repetition and remembering. Paper presented as presidential address before Division of Experimental Psychology, Meetings of the American Psychological Association, Washington, D.C., September 1969.

RUMELHART, D. E. The effects of interpresentation interval on performance in a continuous paired-associate task. Technical Report No. 116, 1967, Institute for Mathematical Studies in the Social Sciences, Stanford University.

WICKELGREN, W. A. Multitrace strength theory. In D. A. Norman (Ed.), Models of memory. New York: Academic Press, 1970.