420

SECTION II

5

RETRIEVAL AS A MEMORY MODIFIER: AN INTERPRETATION OF NEGATIVE RECENCY AND RELATED PHENOMENA

Robert A. Bjork[1]
University of Michigan

Although it is commonplace to assume that the type or level of processing during the input of a verbal item determines the representation of that item in memory, which in turn influences later attempts to store, recognize, or recall that item or similar items, it is much less common to assume that the way in which an item is retrieved from memory is also a potent determiner of that item's subsequent representation in memory. Retrieval from memory is often assumed, implicitly or explicitly, as a process analogous to the way in which the contents of a memory location in a computer are read out, that is, as a process that does not, by itself, modify the state of the retrieved item in memory. In my opinion, however, there is ample evidence for a kind of Heisenberg principle with respect to retrieval processes: an item can seldom, if ever, be retrieved from memory without modifying the representation of that item in memory in significant ways.

It is both appropriate and productive, I think, to analyze retrieval processes within the same kind of levels-of-processing framework formulated by Craik and Lockhart (1972) with respect to input processes; this chapter is an attempt to do so. In the first of the two main sections below, I explore the extent to which negative-recency phenomena in the long-term recall of a list of items is attributable to differences in levels of retrieval during initial recall. In the second section I present some recent results from ex-

[1] Now at the University of California, Los Angeles.

periments designed to assess the differential long-term effects of certain direct manipulations of the way in which an item is initially retrieved.

NEGATIVE RECENCY

In the last six or seven years, there have been a great number of experimental demonstrations that the items at the end of a list, although better recalled initially than any other list items, are the worst-recalled items on tests of final recall administered after substantial delays. The term *negative recency* (Craik, 1970) has come to denote this phenomenon, and I will follow that usage even though I consider the choice of that term to be somewhat unfortunate ("negative recency" has an earlier, different, and generally well-known meaning).

Several Illustrative Experiments

Bjork (1968), Experiment I. In the first of two experiments, Bjork demonstrated negative recency across learning trials in the acquisition of the items in a single free-recall list. Sixty subjects were each presented a list of 40 words three times. The words in the list were presented at a 2-sec rate, and there was a test of immediate free recall following each presentation of the list. The words in the list were scrambled from presentation to presentation, except for eight critical words, two of which were assigned to each of the following four serial-position sequences: MMM, RRM, MMR, and RRR (where M denotes the middle 12 input serial positions in a list presentation, and R denotes the last six input serial positions).

The results of Bjork's first experiment are shown in Fig. 1, in which are two comparisons of particular interest: performance on MMM words versus performance on RRM words, and performance on MMR words versus performance on RRR words. In both comparisons, the words differ in their input positions during the first and second presentations, but they were in the same input portion of the list during the third presentation of the list. Not only were RRM and RRR words not learned better than MMM and MMR words, respectively, as measured by performance following the third presentation, but also they appear to have been learned less well in spite of their greater likelihood of recall following the first and second presentations of the list.

Bjork (1968), Experiment II. In a second experiment, Bjork obtained a similar result. Each of 24 subjects was shown each of five 24-word lists twice. The words in any one list were presented at a 2-sec rate, and an immediate free-recall test followed each presentation of a list. The 24 words in any one list were randomized from the first to the second presentation, except for eight critical words, two of which were Primacy words (presented in one of the first four input positions on both presentations), four of which were Middle words (presented in one of the middle eight input serial positions on both presentations), and two of which were Recency words (presented in one of the last four input positions on both presentations). After the second presentation and recall of the fifth and final list, and after a phony debriefing period lasting several minutes, there was a final free-recall test of subject's memory for all words from all lists.

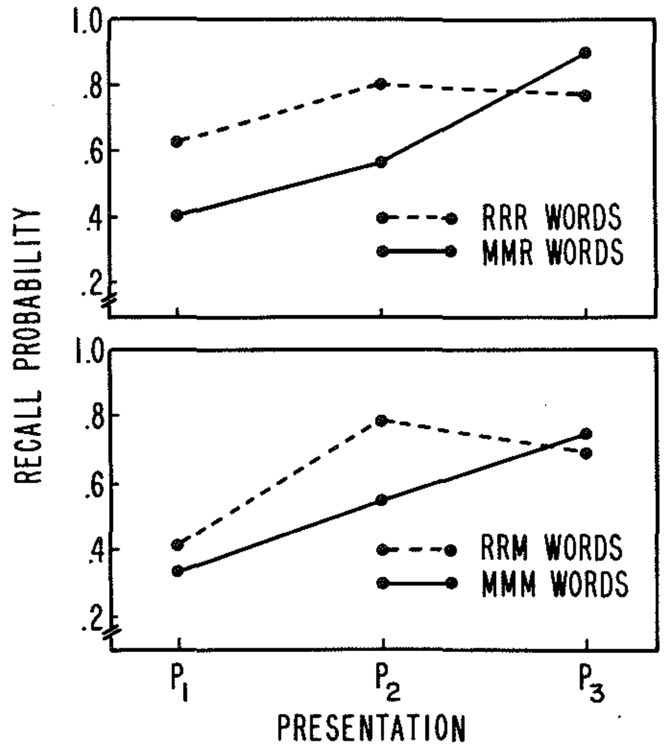

FIG. 1. Learning curves as a function of certain critical sequences of successive input regions across three presentations of a list; R and M denote the recency and middle portions of a list, respectively. (After Bjork, 1968, Experiment I.)

In Table 1 the recall proportions for Primacy, Middle, and Recency words are shown as a function of time of test. Once again, the striking feature of the results in Table 1 is that Recency words, in spite of their higher level of immediate recall following each of the two list presentations, are recalled less frequently on the final test than are Middle words.

TABLE 1
Proportions of Critical Words Recalled Initially and Finally[a]

	Initial tests		
Item type	First	Second	Final test
Primacy items	.50	.70	.54
Middle items	.33	.52	.42
Recency items	.60	.68	.35

[a] After Bjork (1968, Experiment II).

Craik (1970). In an experiment similar to but simpler than the one just reported, Craik obtained a very marked contrast between the initial and final recall of the end items in a list. Each of 80 subjects was presented 10 15-word lists. After each list there was an immediate recall, and after all ten lists there was a final-recall test for all items from all lists.

Craik's results are shown in Fig. 2. The initial-recall curve exhibits the typical strong positive effect of recency, with the recall probability for the very last item in the list approaching one, but the final-recall curve in Fig. 2 decreases systematically across the recency portion of the list.

Madigan and McCabe (1971). With a paired-associate probe procedure, Madigan and McCabe obtained an even more stunning contrast between the initial and final recall of recency items. Each of 30 subjects was presented 50 five-pair lists of paired associates. After each list, the stimulus member of one of the pairs was presented as an immediate probe test of the subject's memory for the paired response. At the end of the experiment there was a final probe retest of all 50 tested pairs.

The results obtained by Madigan and McCabe are shown in Fig. 3. Recall of the response member of the fifth pair in a list was all but perfect initially, but those same response members were never recalled on the final test.

A Depth-of-Retrieval Interpretation of Negative Recency

One reason that negative-recency phenomena have aroused considerable interest is that they are unintuitive. If one simply views retrieval as an important learning event, the conditions that maximize initial recall should, in turn, maximize final recall. From that standpoint, the last item in a list should be the best-recalled item in final recall rather than the worst-recalled item.

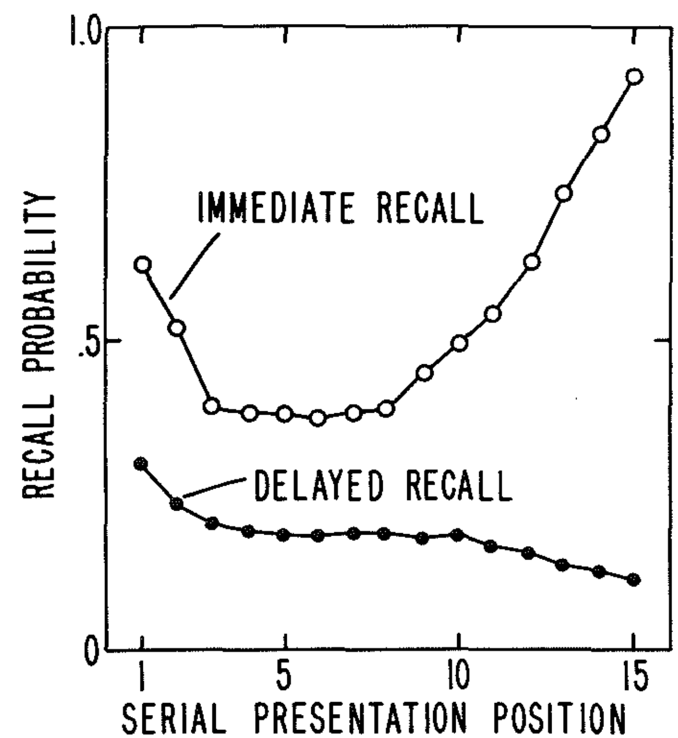

FIG. 2. Immediate and final recall proportions as a function of input serial position. (After Craik, 1970.)

It is unrealistic, however, to assume that retrieval from memory is a singular event that has uniform consequences on the state of the retrieved item. One interpretation of negative recency (Bjork, 1970; Craik, 1970) is based on the assumption that the long-term benefits of an initial retrieval are an increasing function of the depth or difficulty of the initial retrieval. In particular, retrieval from short-term memory (STM) is assumed to consist of a kind of rapid dumping, which has little, if any, effect on later efforts to retrieve those items from long-term memory (LTM), whereas an initial retrieval that itself constitutes a retrieval from LTM does facilitate later efforts to retrieve from long-term memory. Thus, negative recency comes about because of differential effects of the initial retrieval: long-term recall of items in the beginning and middle of a list is enhanced by an initial retrieval, whereas long-term retrieval of items at the end of a list does not profit from their initial retrieval from STM.

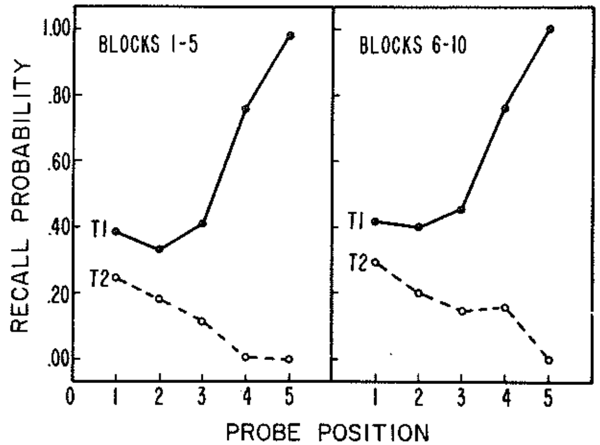

FIG. 3. Immediate (T1) and final (T2) paired-associate recall probabilities as a function of input position. (After Madigan & McCabe, 1971.)

In order to test the depth-of-retrieval interpretation of negative recency, I designed an experiment (Bjork, 1970, Experiment I) similar to Craik's (1970) experiment reported above, except that the initial presentation of a list either was or was not followed by an initial test. Each of 32 subjects was presented eight 16-word lists. After each list there was a 30-sec period during which subjects either recalled the list or shadowed nine-digit numbers presented at a 2-sec rate. Four of the lists were followed by an initial recall, and the other four lists were followed by digit shadowing. Subjects could not anticipate whether a given list would be followed by an initial recall or digit shadowing because that activity was cued by a postinput cue (a row of question marks or a nine-digit number presented after the last word in a list), and the cuing was haphazard across the eight lists. At the end of the experiment there was a final free-recall test for all items from all lists.

The obtained results (shown in Fig. 4) are quite consistent with the depth-of-retrieval interpretation of negative recency. Overall, final recall was enhanced by the initial recall, but the item that profited least from the initial recall was that item recalled most frequently during the initial recall—the last item in the list.

An Amount-of-Rehearsal Interpretation of Negative Recency

Negative recency also may be interpreted as reflecting differences in amount of rehearsal prior to an initial retrieval. In this view, recency items

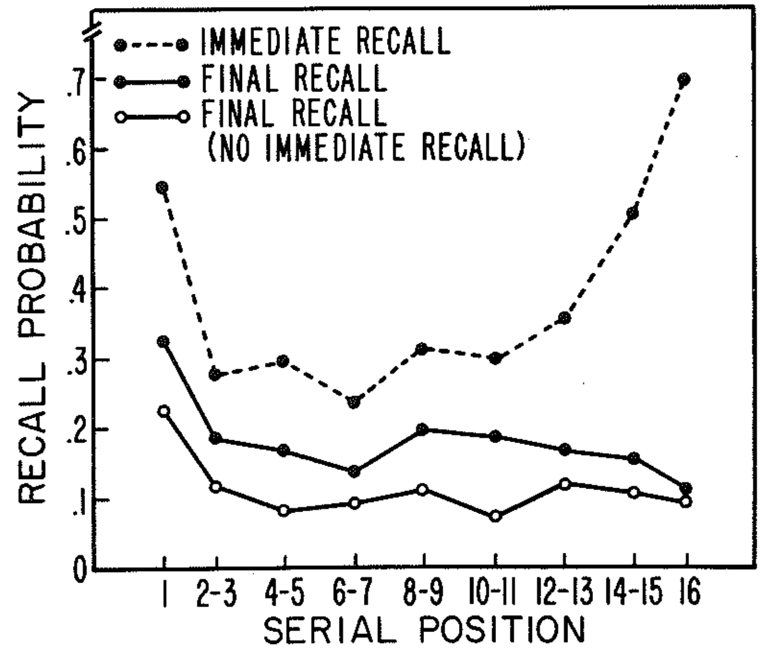

FIG. 4. Immediate and final free recall as a function of input serial position. (After Bjork, 1970, Experiment I.)

receive less rehearsal prior to initial recall that do earlier items in the list. The near-perfect initial recall of recency items is attributable to the strong likelihood that those items remain available in STM at the time of the initial recall. In terms of their strength of representation in LTM, however, recency items are weaker than any other items, and that weakness results in their inferior long-term recall. Thus, no appeal to differential effects at the time of retrieval is necessary to account for negative-recency phenomena.

One problem with the amount-of-rehearsal interpretation is that it is inconsistent with the results shown in Fig. 4. If initial retrieval were assumed to have no effect on later retrieval, then the two final-recall curves should fall on top of each other. If initial retrieval were assumed to have a uniform positive effect on later retrieval, then the two final-recall curves in Fig. 4 should diverge rather than converge across the recency portion of the list, since those items are recalled more frequently than are any other items in the initial-recall lists.

That problem notwithstanding, Rundus and Atkinson (1970) obtained what seemed to be strong evidence in favor of the amount-of-rehearsal interpretation of negative recency. Subjects were presented a series of 20-word lists after each of which there was a test of immediate free recall, and after all of which there was a test of final free recall. The words in any one list were presented at a relatively slow rate (5 sec per word). Subjects were required to rehearse aloud, and their overt rehearsal of the words in a list was recorded. Rundus and Atkinson obtained the typical positive and negative effects of recency in the initial and final tests of free recall, respectively. The result of particular interest, however, was that there was a negative-recency effect in the number of overt rehearsals devoted to an item as a function of its input serial position. That is, the last item or two in a list did, in fact, receive fewer overt rehearsals prior to immediate recall than did earlier items in the list.

Although the Rundus and Atkinson results appear to provide strong support for an amount-of-rehearsal interpretation of negative recency, other considerations and evidence render the amount-of-rehearsal interpretation altogether untenable. One problem is that the words actually recalled initially should, from the standpoint of the final recall, profit at least as much from their overt recall as they do from any one within-list overt rehearsal. When the observed likelihood of initial recall as a function of input position is added to the overt-rehearsals function, the resulting function exhibits little if any negative recency. Any argument about the extent to which that problem is a serious problem becomes academic, however, in view of results obtained by Craik and Watkins (1973, Experiment II) and by Light (1974).

In the Craik and Watkins experiment, 16 subjects were each presented 12 lists with 12 words in each list. The words in any one list were presented at a 3-sec rate, subjects were required to rehearse aloud, and each subject's overt rehearsals were recorded. The last four words in each list were printed in capital letters, and the subjects were instructed to recall those four items first during the initial recall that followed each list. For six of the 12 lists the initial recall was immediate, and for the other six lists the initial recall was delayed by a 20-sec unfilled period, during which subjects were free to continue their overt rehearsal of the words in that list. At the end of the experiment, there was a test of final free recall.

The results of the Craik and Watkins experiment are shown in Fig. 5. Even though the last four list items receive a great deal more rehearsal prior to the delayed initial free recall than they do prior to the immediate initial free recall, the level of final free recall is essentially the same in both cases, and there is negative recency in the final-recall curve in both cases.

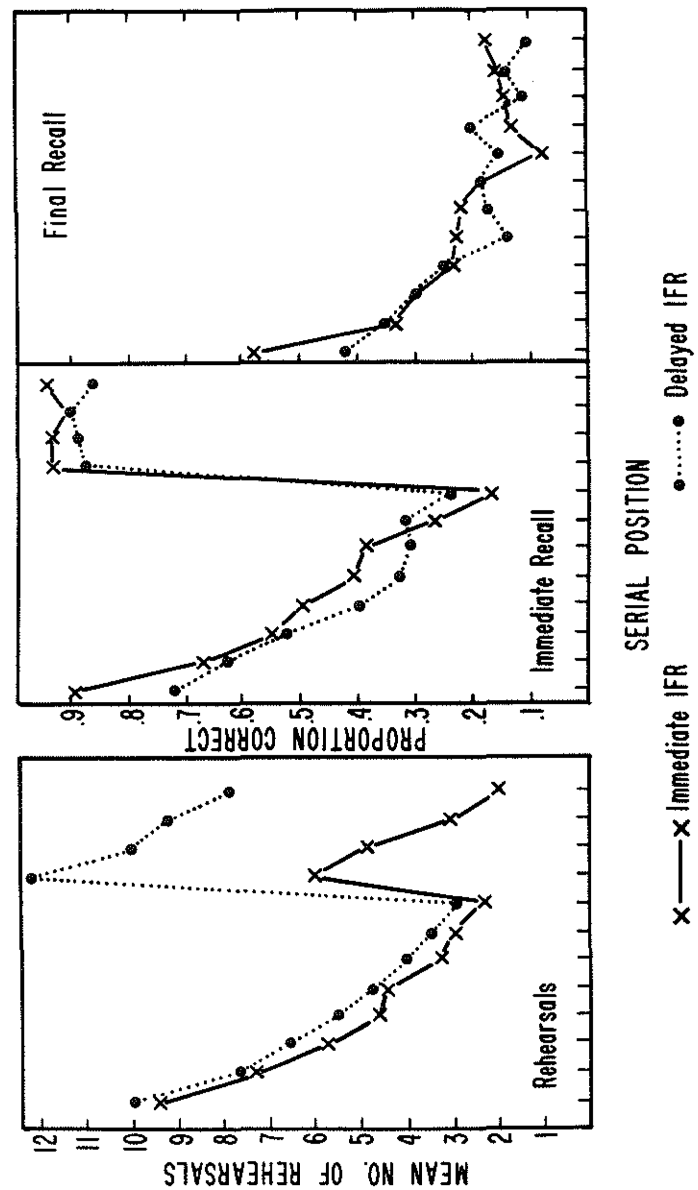

FIG. 5. Number of (a) overt rehearsals and probabilities of (b) initial and (c) final free recall as a function of input serial position and delay of initial recall: (×) immediate IFR; (●) delayed IFR. (After Craik & Watkins, 1973.)

By themselves, the Craik and Watkins results indicate quite dramatically that negative recency is not attributable to a deficiency in the *amount* of rehearsal given recency items prior to initial recall, but the results of two recent experiments by Light (1974) snuff out any lingering hope one might have for the amount-of-rehearsal interpretation of negative recency. Using a paired-associate probe paradigm similar to that employed by Madigan and McCabe (1971), Light increased both the incentive and the time available for subjects to rehearse the terminal items in a list. In spite of the additional rehearsal of terminal pairs effected by those manipulations, there remained sizable negative-recency effects in final probed recall.

A Levels-of-Processing Interpretation of Negative Recency

Although there is little or no evidence for the amount-of-rehearsal interpretation of negative recency, there is substantial evidence that differences in the *type* or *level* of processing of list items prior to an initial recall are at least as important as differences in the depth of initial retrieval in producing negative recency in long-term recall. As has been demonstrated by a number of recent studies (Bjork & Jongeward, 1974; Craik & Watkins, 1973; Jacoby, 1973; Jacoby & Bartz, 1972; Mazuryk, 1974; Mazuryk & Lockhart, 1974; Meunier, Ritz, & Meunier, 1972; Woodward, Bjork, & Jongeward, 1973), it is necessary to distinguish between rehearsal as a rote, cyclic activity and rehearsal as a constructive associative activity. The former type of rehearsal, referred to as *primary* rehearsal (Woodward *et al.*, 1973) or *maintenance* rehearsal (Craik & Watkins, 1973), does not facilitate long-term recall, whereas the latter type of rehearsal, referred to as *secondary* rehearsal (Woodward *et al.*, 1973) or *elaborative* rehearsal (Craik & Watkins, 1973), does facilitate long-term recall. The position I would like to argue in this section is that the terminal items in a list receive less secondary rehearsal or processing prior to initial recall than do earlier items, and that deficiency, together with their more superficial retrieval during initial recall, results in their inferior long-term retention.

Before I cite some supporting evidence for the levels-of-processing interpretation of negative recency, I want to specify the position in more detail. As a prototypical situation, consider the processing devoted to a single word in a typical free-recall list. During the time a word is presented, a subject may engage in some mixture of primary and secondary processing activities. Those activities may vary substantially from one word to the next, but the nature of such processing should not vary systematically with a word's input serial position. The initial few words in a list might receive more processing during their presentations because of the minimal competition for processing time from earlier items in the list, and that greater processing might in fact produce the positive primacy observed in both initial and final recall, but after the first few items, the processing during an item's presentation should be relatively uniform across input serial position. A similar pattern should hold for the processing of an item during the brief period just subsequent to its presentation, and any such rehearsal will tend to be of the primary or maintenance type, which means that it will have little consequence for long-term recall in any case. As the interval from the presentation of a particular item lengthens, however, and as more subsequent items are presented, any within-list retrievals (rehearsals) of that item will be increasingly of the secondary or elaborative type. The likelihood that a given item will be retrieved or rehearsed will decrease, of course, as the interval from its presentation increases, but any within-list retrievals that do occur will be increasingly potent in terms of their effect on long-term recall. According to the present characterization, then, the inferior long-term recall of recency items is attributable to two consequences of their being last presented and first recalled: (*a*) their immediate superficial readout from STM does not facilitate their later retrieval from LTM, and (*b*) compared to earlier list items, the opportunities for within-list retrievals or rehearsal of the secondary type are severely restricted. I do not, of course, mean to imply that these assumed differences in processing during input and output are independent of each other; clearly, any assumed differences in retrieval processes during output must be in large part determined by differences in encoding, rehearsal, and forgetting processes during input.

No additional support for the first of the two assumed components of negative recency will be cited here. Some evidence that depth-of-retrieval during initial recall is a factor in negative recency has already been cited, and the phenomena discussed in the next section constitute stronger if less direct evidence for such an assumption.

Evidence for the contribution of within-list processing to negative recency derives from several sources. First, Mazuryk (1974) has shown that it is possible to produce positive rather than negative recency in final free recall if subjects are cued to rehearse the recency items in a list in an associative (or secondary) fashion. In Mazuryk's experiment, subjects were presented 12 14-word lists and immediate recall was required after each list. During the presentation of the tenth word in each list, subjects were instructed, by means of prearranged cues, how to process the last four words in the list. They were cued either to rehearse each of the last four items in turn in a rote (primary) fashion—either overtly or covertly—or they were cued to generate as many verbal associates to each of the last four words as they could. The two rote-rehearsal conditions (covert and overt) produced marked positive recency in immediate recall (recall of the four recency items was near-perfect), but those conditions also resulted in negative recency in final recall. The associate-processing condition, in contrast, resulted in positive recency in both immediate and final recall (the levels of initial and final recall of the four recency items were lower and much higher, respectively, than were the levels of initial and final recall of those items in the rote-rehearsal conditions).

Mazuryk's results are consistent with the notion that recency items are typically processed in a superficial fashion and that their early recall tends to truncate the opportunity for deeper postpresentation processing of those items. That interpretation is also supported in a quite different way by the results of an experiment by Glenberg and Melton (1974). In Glenberg and Melton's experiment, subjects were required to rehearse aloud during the presentation of several free-recall lists, each of which was followed by an immediate free recall. Of the various analyses of subjects' rehearsal processes carried out by Glenberg and Melton, the analysis of particular interest involves items in the middle rather than at the end of a given list. For items whose free recall was not influenced by primacy or recency, Glenberg and Melton were clever enough to look at recall probability as a function of not only the number of overt rehearsals but also as a function of the spacing of those rehearsals. That is, if an item was given a certain fixed number of overt rehearsals, recall was plotted as a function of the average spacing between successive rehearsals.

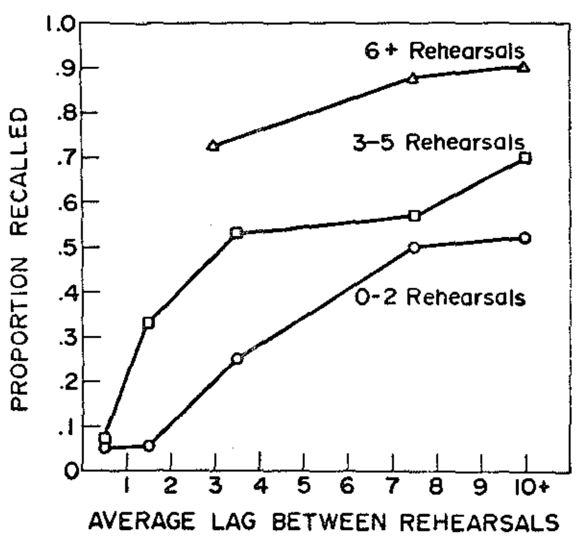

FIG. 6. Free-recall probabilities as a function of the average spacing of a word's within-list overt rehearsals. (After Glenberg & Melton, 1974.)

Glenberg and Melton's results are shown in Fig. 6. Given a fixed number of rehearsals, there were quite remarkable increases in recall as a function of the spacing of those rehearsals.

The Glenberg and Melton results are correlational in nature, which means that there are at least two different ways of looking at those results. The least interesting and least plausible interpretation, from my standpoint, is the possibility that the increase in recall as a function of spacing has nothing to do with spacing per se, but that, rather, both spacing of overt rehearsals and initial free recall increase as a function of the degree to which items are well learned when they are initially presented. Thus, if items are not reasonably well learned when they are presented initially, they become unavailable for rehearsal quickly, which means that they will not survive long interrehearsal intervals, whereas well-learned items will survive such intervals. Although the results in Fig. 6 might reflect nothing more than such a selection effect, it seems more plausible to me that they reflect primarily the increased long-term benefits of a within-list retrieval as a function of the delay of that retrieval from the preceding retrieval or input of that item.

The level-of-processing interpretation of negative recency put forward in this section views negative recency phenomena as the natural consequences of a subject's rehearsal and retrieval processes during the input and output, respectively, of a list of items. Götz and Jacoby (1974) and Mazuryk and Lockhart (1974) have also interpreted negative recency phenomena in terms of a kind of level of processing framework, but their interpretations differ somewhat from the present view. In both the Götz and Jacoby (1974) and Mazuryk and Lockhart (1974) interpretations, retrieval processes during initial recall are not, by themselves, assumed to play a major role in producing negative recency during final recall. Rather, the explanatory burden is put on differences as a function of serial position in the level of processing (or type of retrieval cues established) during the input of a list. Such differences are not assumed to reflect natural or typical processing activities as much as they are assumed to reflect a subject's efforts to maximize the level of initial recall. The subject is assumed, in anticipation of the end of a list and the subsequent test of immediate recall, to intentionally encode the last few list items in a low-level fashion that makes them readily available for initial recall but impairs their long-term recall. Such a view requires, of course, that subjects be able to anticipate the end of a list. One testable implication of the Götz and Jacoby (1974) and Mazuryk and Lockhart (1974) interpretations is that list-length uncertainty (that is, introducing substantial variation in the number of items presented in each of a series of lists presented for initial recall) should reduce greatly or abolish negative-recency effects in final recall.

RELATED PHENOMENA

There is little doubt, at least in my own mind, that negative-recency phenomena reflect some very important differences in the extent to which different types of retrieval processes modify the long-term state of an item in memory. From a procedural standpoint, however, the negative-recency paradigm constitutes a somewhat indirect means of manipulating retrieval processes. In the present section, the long-term consequences of some more direct manipulations of the delay, difficulty, and type of initial retrieval are reviewed.

Delay of Retrieval

In order to examine more directly the extent to which the delay of an initial retrieval influences the effectiveness of that retrieval as a memory modifier, Whitten and Bjork (1972) devised a paradigm that is a kind of hybrid combination of the standard free-recall and Brown–Peterson paradigms. Subjects were each shown several lists of words to free recall, but the lists differed from typical free-recall lists in several important ways: (*a*) the 24 words in a list were presented as 12 word doubles rather than as single words; (*b*) the presentation of any two successive word doubles was separated by 22-sec of intervening distractor activity; (*c*) embedded at one of three different points within the distractor activity following a given word double (after retention intervals of 4, 8, or 14 sec), there was a 3-sec interval during which the word double was either presented again or was tested for recall; and (*d*) the free-recall test following any one list was delayed by a period of distractor activity long enough to nullify the influence of short-term memory.

The results of Whitten and Bjork's (1972) experiment are shown in Fig. 7. The proportions of words correctly recalled on the embedded within-list tests and on the free-recall test following a given list are plotted as a function of the delay from the initial presentation of a word double until the within-list test or second presentation of that word double. The top two curves in Fig. 7 are simply what one might expect: The fact that performance on the within-list test decreases with the delay of that test is hardly surprising, and the increase in the free recall of words in the present–present condition as a function of spacing is a result that has also become altogether commonplace.

It is the bottom curve in Fig. 7 that is of particular interest. The fact that performance in the present-test condition increases with the delay of the initial test is by no means an obvious result. If one simply views correct initial retrieval as a learning event, then one would expect the probability

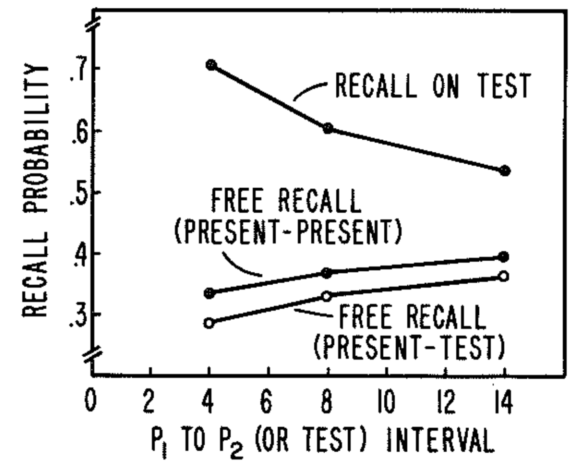

FIG. 7. Within-list recall and subsequent free recall as a function of presentation–presentation or presentation–test interval. (After Whitten & Bjork, 1972.)

of subsequent retrieval to decrease rather than increase with the delay of the initial retrieval. If, however, one assumes that the extent to which an act of retrieval serves as a learning event is a function of the depth, difficulty, or complexity of that act, then the results of the present-test condition are not surprising. Landauer and Eldridge (1967) obtained a somewhat similar result using a paired-associate paradigm, although the increase in performance as a function of presentation–test interval in their situation seemed restricted to very short intervals.

Difficulty of Retrieval

In order to interpret Whitten and Bjork's (1972) results as supporting the notion that the effectiveness of a retrieval act is an increasing function of the difficulty of that act, it is necessary to assume retrieval difficulty increases with retrieval delay. Although such an assumption seems altogether reasonable, the results of an experiment by Gardiner, Craik, and Bleasdale (1973), and the results of several unpublished experiments carried out by Eva Bradford, illustrate in a more direct fashion that the difficulty or complexity of an initial retrieval influences its effect on subsequent retrieval.

In the Gardiner *et al.* (1973) experiment, the definition and first letter of each of 50 words (frequency less than one per million) were presented to subjects who were instructed to retrieve that word as quickly as possible.

The retrieval latency for each word that was, in fact, retrieved was recorded, and, at the end of the experiment, subjects were asked to write down as many of the 50 words as they could remember. In general, final-recall probability increased with initial-retrieval latency: only 27% of those words retrieved within 15 sec were recalled on the final test, whereas 48% of those words retrieved between 15 and 60 sec were recalled.

One problem with the Gardiner *et al.* (1973) results is that the differences in final-recall probability as a function of initial-retrieval latency might be attributable to the amount of time a subject is exposed to the definition and first letter of a given word, rather than to the difficulty of the initial retrieval. That is, during the test of final recall, subjects' memory for definitions and first letters presented earlier might mediate their recall of the corresponding words.

Such mediation problems are avoided somewhat, but not entirely, by the design employed by Eva Bradford in several unpublished experiments. On each of a series of trials in Bradford's experiments, subjects were instructed to give an associate of a certain type in response to the presentation of a word. In the simplest of her experiments, one of two prearranged cues was presented prior to the presentation of a given word. One cue instructed subjects to give their first associate to the subsequent word, whereas the other cue instructed them to give a novel associate, that is, an associate that they thought other people would be very unlikely to give in response to that word. At the end of the experiment, subjects were asked, without forewarning, to recall as many as possible of the associates they had generated during the experimental session.

Bradford's results are consistent with the notion that the extent to which a later recall profits from an initial recall is a function of the difficulty or complexity of the retrieval processes during the initial recall. In the simple experiment described above, 45% of the novel associates were recalled on the final test, whereas only 31% of the first associates were recalled. In a more complex experiment in which subjects were cued to give second associates, superordinates, coordinates, and so forth, as well as first associates and novel associates, there was also a general positive correlation between the level of final recall and the difficulty of initial retrieval.

Although it seems likely in Bradford's experiments that the recall of associates during final recall was mediated, in part, by subjects' memory for the presented words that gave rise to those associates during the experimental session, it seems unlikely that the direction of the differences in final recall could be attributable to such processes. One would expect, for example, that it would be easier to regenerate and recognize having given a first associate to a given word than it would be to regenerate and recognize a second or novel associate.

Levels and Stages of Retrieval

A recent dissertation by Whitten (1974) constitutes what is probably the most tightly controlled and analytical investigation of test events as learning trials. In each of several experiments, Whitten employed postinput cues to subjects as a means of manipulating the nature of an initial retrieval without also influencing the nature of the encoding processes during the initial presentation of an item. It is beyond the scope of this chapter to present either the method or the results of Whitten's several experiments in full detail, but the following incomplete descriptions of two of his experiments should clarify the general methodology and some of the principal results of his research.

The structure of a typical trial in the first of Whitten's experiments is diagrammed in Fig. 8. After a READY signal and a 1-sec blank period, each of four common words was presented for .75 sec in the window of a high-speed memory drum. A .50-sec search-mode signal followed the fourth word on a trial and preceded the presentation of a probe (test) word. The search-mode signal (RRRRRR or MMMMMM) served as a postinput cue to subjects to recall aloud the list word that either rhymed with (RRRRRR) or meant the same as (MMMMMM) the probe word. In the particular illustration shown in Fig. 8, the correct response on an RRRRRR trial would be the list word rhyming with STATION ("NATION"), and the correct response on an MMMMMM trial would be the list word that meant the same as COUNTRY (also "NATION"). The lists were constructed such that, across subjects, the same pool of words was recalled on both acoustic-search and semantic-search trials. At the end of the ex-

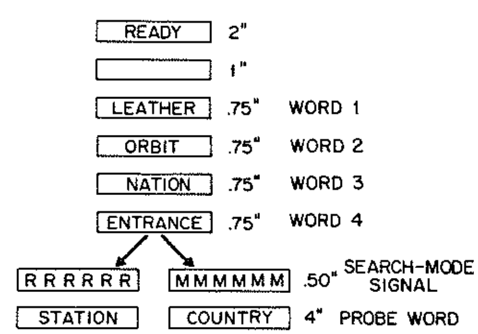

FIG. 8. Sequences of events on a typical acoustic-search (left arrow) or semantic-search (right arrow) trial in Whitten's (1974) Experiment I.

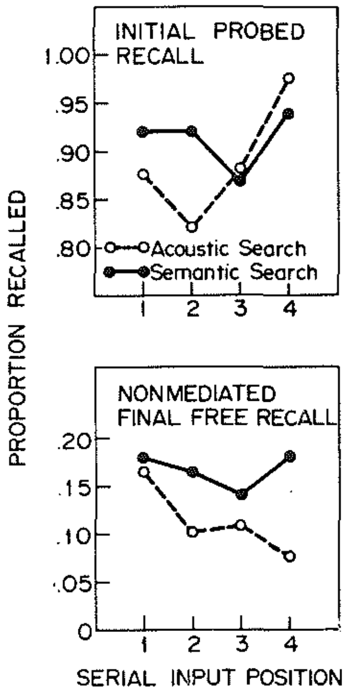

FIG. 9. Initial and final recall as a function of input serial position and initial-test search mode. (After Whitten, 1974.)

periment, subjects were asked to recall as many of the words presented during the experiment as they could.

The proportion of tested words recalled initially on acoustic-search and semantic-search trials is shown as a function of serial input position in Fig. 9, together with the nonmediated final-recall proportions for those words. "Nonmediated final free recall" denotes a scoring procedure in which cases where the recall of a probed list word was preceded and possibly mediated by the just prior recall of its corresponding probe word were excluded from the analysis. Whether such cases were or were not excluded did not, in fact, change the basic pattern of results on the final recall.

Several aspects of Whitten's results in Fig. 9 merit comment:

1. The initial test of a word, whether based on an acoustic search or semantic search, contributed greatly to its final recall. The final-recall proportions for list words that were not tested initially fell between .02 and .04 in all conditions.

2. The initial-recall curve in the acoustic-search condition exhibits positive recency, whereas the final-recall curve exhibits negative recency. If one assumes that the acoustic code of a word is lost rapidly enough that only the fourth word in a list was reliably retrieved on the basis of its acoustic trace, and that the earlier words in a list—the first word in particular—tended to be retrieved on the basis of a deeper-level encoding of some type, then the final-recall curve can be interpreted as demonstrating that initial retrieval on the basis of an acoustic trace does relatively little to facilitate later recall.

3. Initial retrieval on the basis of a semantic search, however, appears both to be less sensitive to input recency and to have much more substantial effects on final recall.

In his research on tests as learning trials, Whitten was concerned not only with "depth" of retrieval as a factor, but also with the relative contributions of the search and recovery components of a retrieval process. Thus, if a set of items is presented and there is an initial effort to retrieve one of those items based on some cue, any later attempt to retrieve those items will reflect encoding processes during their presentation plus whatever modifications in the state of those items are attributable to searching through the set and recovering (recalling) a given item at the time of the initial retrieval.

In order to gain some information about the relative contributions of initial search and initial recovery to final recall, and in order to gain information on several other questions as well, Whitten devised a very clever and very complex experiment. The procedure was much the same as that outlined in Fig. 8 except that (*a*) two words rather than four words were presented on a given trial; (*b*) on some trials neither of the two list words matched the probe word in terms of the cued relation, although in some of such cases there was a match between one of the list words and the probe word in terms of the uncued relation (the correct response on such trials, whether the uncued relation existed or not, was "no match"); (*c*) on some trials there was no test at all (on such trials, NO TEST replaced the search-mode signal and a simple arithmetic problem replaced the probe word); and (*d*) on trials where the cued relation did exist between one of the list words and the probe word, the uncued relation might or might not exist between the probe and the other list word. In addition to obtaining the probabilities of initial probed recall and final free recall, Whitten also obtained the latencies of initial recall and the probabilities of final recognition. The final-free-recall proportions shown in Fig. 10 constitute a small subset of the results from Whitten's experiment (for the full details, see Whitten, 1974).

In the condition designations at the bottom of Fig. 10, the first two letters indicate the relation between each of the two list words and the probe word (A = acoustic, S = semantic, and X = unrelated), and the letters

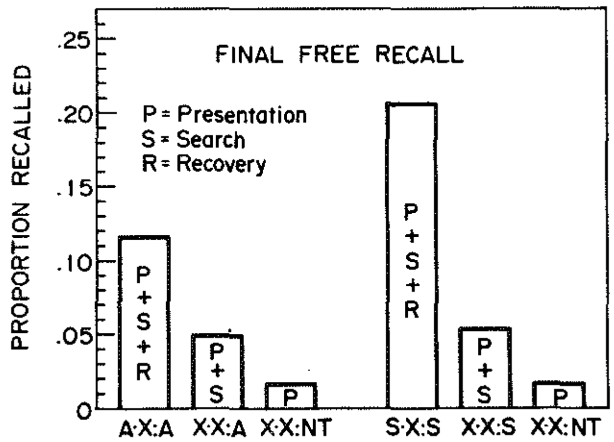

FIG. 10. Final free recall as a function of initial list type, where P is presentation, S is search, and R is recovery. (After Whitten, 1974, Experiment III.)

after the colons indicate the cued search mode (A = acoustic, S = semantic, and NT = no test). Several aspects of the results in Fig. 10 merit comment:

1. Words that are presented and initially recalled are much better recalled finally than are words not initially recalled.
2. Initial retrieval of a list word based on its semantic properties produces substantially better final recall of that word than does initial retrieval of a word based on its acoustic properties.
3. Finally, searching through a set of items in memory to see whether any of those items have a particular acoustic or semantic property appears to facilitate the later recall of those items, although the extent of such facilitation is small compared to cases where an item is actually recovered (recalled) as an outcome of an initial search.

CONCLUSION

The phenomena reviewed in this chapter constitute compelling evidence that an item's state in memory is modified by its retrieval and, more importantly, that the extent of such modification is a function of the depth or level of the retrieval processes involved. There are several thoretical and practical implications of this general result that merit comment:

1. There is one sense in which there is nothing surprising in the present results. Given that level of processing during the input of a verbal item has been shown to be a potent factor in determining one's retention of

that item, and that the degree to which rehearsal facilitates later recall is also a function of the level of processing involved, there seems little reason not to expect a similar relationship in the case of retrieval processes.

2. Although the present results seem consistent with other recent results, it remains a nontrivial matter to say why the effectiveness of test trials as learning events should depend on the level or depth of the retrieval processes involved. My own guess is that deeper, more difficult, more complex retrieval processes have two distinct long-term advantages over shallower, less difficult, and less complex retrieval processes: (*a*) they reactivate or strengthen encodings of an item that are more durable, less susceptible to interference, and, therefore, more supportive of long-term retention, and (*b*) because they involve taking a slower and more complex route to an item in memory, they multiply or elaborate the routes available for subsequent retrieval. The fact that long-term recognition seems to profit less from an initial retrieval than does long-term recall (see, e.g., Hogan & Kintsch, 1971; Whitten, 1974) might reflect the relative unimportance of the second factor mentioned above when long-term recognition rather than long-term recall is tested.

3. Whatever the mechanisms involved, the present results have some obvious practical implications. Tests have long been appreciated as important learning events in educational contexts, but there has been some tendency to structure the tests embodied in a learning program in such a way that the typical student will seldom if ever fail to come up with the correct answer. In part, such trivial tests are apparently motivated by the notion that the production of a correct response optimizes the learning on a test. The present results, to the extent that they can be generalized to an educational context, suggest that efforts in programmed texts and elsewhere to optimize the learning process by means of frequent superficial tests designed to produce near-perfect reponding are misguided.

ACKNOWLEDGMENTS

The preparation of this chapter and the author's research reported herein were supported by the Advanced Research Projects Agency, monitored by the Air Force Office of Scientific Research under Contracts F44620-72-C-0019 and AF49(638)-1235, respectively, with the Human Performance Center, Department of Psychology, University of Michigan.

REFERENCES

Bjork, R. A. The short-term and long-term effects of recency in free recall. Paper presented at the Meeting of the Psychonomic Society, St. Louis, Missouri, 1968.
Bjork, R. A. Control processes and serial position effects in free recall. Paper presented at the Mathematical Psychology Meetings, Miami, Florida, 1970.

Bjork, R. A., & Jongeward, R. H., Jr. Rehearsal and mere rehearsal. Unpublished manuscript, 1974.

Craik, F. I. M. Fate of primary memory items in free recall. *Journal of Verbal Learning & Verbal Behavior,* 1970, **9,** 143–148.

Craik, F. I. M., & Lockhart, R. S. Levels of processing: A framework for memory research. *Journal of Verbal Learning & Verbal Behavior,* 1972, **11,** 671–684.

Craik, F. I. M., & Watkins, M. J. The role of rehearsal in short-term memory. *Journal of Verbal Learning & Verbal Behavior,* 1973, **12,** 599–607.

Gardiner, J. M., Craik, F. I. M., & Bleasdale, F. A. Retrieval difficulty and subsequent recall. *Memory & Cognition,* 1973, **1,** 213–216.

Glenberg, A. M., & Melton, A. W. Rehearsal processes associated with degree of spacing of repeated items in free recall learning. Unpublished manuscript, 1974.

Götz, A., & Jacoby, L. Encoding and retrieval processes in long-term retention. *Journal of Experimental Psychology,* 1974, **102,** 291–297.

Hogan, R. M., & Kintsch, W. Differential effects of study and test trials on long-term recognition and recall. *Journal of Verbal Learning & Verbal Behavior,* 1971, **10,** 562–567.

Jacoby, L. L. Encoding processes, rehearsal, and recall requirements. *Journal of Verbal Learning & Verbal Behavior,* 1973, **12,** 302–310.

Jacoby, L. L., & Bartz, W. A. Rehearsal and transfer to LTM. *Journal of Verbal Learning & Verbal Behavior,* 1972, **11,** 561–565.

Landauer, T. K., & Eldridge, L. Effect of tests without feedback and presentation-test interval in paired-associate learning. *Journal of Experimental Psychology,* 1967, **75,** 290–298.

Light, L. L. Incentives, information, rehearsal, and the negative recency effect. *Memory & Cognition,* 1974, **2,** 295–300.

Madigan, S. A., & McCabe, L. Perfect recall and total forgetting: A problem for models of short-term forgetting. *Journal of Verbal Learning & Verbal Behavior,* 1971, **10,** 101–106.

Mazuryk, G. F. Positive recency in final free recall. *Journal of Experimental Psychology,* 1974, **103,** 812–814.

Mazuryk, G. F., & Lockhart, R. S. Negative recency and levels of processing in free recall. *Canadian Journal of Psychology,* 1974, **28,** 114–123.

Meunier, G. F., Ritz, D., & Meunier, J. A. Rehearsal of individual items in short-term memory. *Journal of Experimental Psychology,* 1972, **95,** 465–467.

Rundus, D., & Atkinson, R. C. Rehearsal processes in free recall: A Procedure for direct observation. *Journal of Verbal Learning & Verbal Behavior,* 1970, **9,** 99–105.

Whitten, W. B. Retrieval "depth" and retrieval component processes: A levels-of-processing interpretation of learning during retrieval. Technical Report No. 54, Human Performance Center, University of Michigan, Ann Arbor, Michigan 1974.

Whitten, W. B., & Bjork, R. A. Test events as learning trials: The importance of being imperfect. Paper presented at the Midwestern Mathematical Psychology Meetings, Bloomington, Indiana, 1972.

Woodward, A. E., Jr., Bjork, R. A., & Jongeward, R. H., Jr. Recall and recognition as a function of primary rehearsal. *Journal of Verbal Learning & Verbal Behavior,* 1973, **12,** 608–617.