Mnemonic Elaboration in Multilist Learning

GORDON H. BOWER

Stanford University, Stanford, California, 94305

AND

JUDITH S. REITMAN

The University of Michigan, Ann Arbor, Michigan, 48104

This study inquires whether retroactive interference (RI) from learning multiple word lists can be altered through mnemonic strategies. The Ss learned five successive lists of 20 words each by associating the words with 20 conceptual pegs via visual imagery. Some Ss were instructed to visualize the words from the successive lists in entirely new associative scenes; other Ss were to incorporate the current list's words into the appropriate scenes from the earlier lists. Although equivalent in immediate recall, the Separate Images (SI) group recalled less than the Progressive Elaboration (PE) group at the end of the session, and showed a strict RI curve across lists where the PE group did not. Recall at a one-week test, however, favored earlier lists for the PE group only, explicable in terms of cumulative rehearsal of earlier items with the PE method.

Numerous recent researches [reviewed by Parvio (1971, p. 332 ff.)] have proven the efficacy of the "pegword mnemonic" system for learning lists of unrelated items. In this procedure, the person first learns a series of concrete words (the "pegs") associated with the first 20 or so integers. Often-cited examples are the rhyming pegs "1 is a bun, 2 is a shoe, 3 is a tree, . . ." and so on, up to 20 or more. To learn any new list of items, the person is instructed to imagine the referent of the nth to-be-remembered word in vivid interaction with the referent of the nth pegword. When asked to recall the list, the S implicitly recites his known pegword list, with each peg cueing recall of its associated item.

Prior research (Bugelski, 1968; Keppel & Zavortink, 1969) has shown that use of the pegwords for learning several lists in succession creates retroactive interference (RI) for recall of the earlier lists. Multilist learning with the pegwords is functionally like the A-B, A-C paradigm of negative transfer. Indeed, appropriate controlled comparisons in our laboratory [see also Wood (1967)] have shown significant negative transfer from use of the same pegs for learning several lists. One of our interests in the experiment to be reported was to develop a simple procedure to reduce or eliminate the RI that accumulates from multiple use of the peglist. We call the technique "progressive elaboration" since it resembles progressive addition of new objects to a picture: when S is presented with the nth item in List J, he is to call up his nth pegword (or image), try to remember all the prior objects from Lists 1, 2, . . . , (J - 1) that he has placed in that imaginal scene, and then add the referent of the new word from List J to this grand scene, trying to concoct some active relation between the new referent image and those objects already in the scene. For

---

1 This research was supported by a research grant, MH-13950-05, to the first author from the National Institute of Mental Health. The same agency provided a post-doctoral fellowship, 1-F02-GM-44, 533-01, to the second author, during which time this research was conducted.
example, suppose the third word in Lists 1, 2, 3 were *swing, cigar, fish*, respectively. With the pegword "3 is a *tree*," the scenes progressively elaborated during learning of the lists might be: for List 1, "a *swing* hanging from branches of a *tree*"; for List 2, "a *cigar* lying on the *swing* on the *tree*"; for List 3, "a *fish* smoking a *cigar* while *swinging* under a *tree*.

In the following experiment, Ss using this progressive-elaboration strategy were compared to control Ss instructed to concoct entirely new imaginal scenes for each list, specifically not calling to mind the words from previous lists associated to each pegword as it was used anew. The expectation is that Ss using the "progressive elaboration" strategy for learning successive lists will show much less RI across successive lists than will the control Ss instructed to use separate images to associate the items from successive list to their corresponding pegs.

In addition to testing this "progressive elaboration" technique, the experiment was designed to answer several further questions about mnemonics. One question was whether the "method of loci" is essentially equivalent in performance characteristics to the pegword mnemonic. In the method of loci, the S learns an ordered list of "mental snapshots" of geographic locations such as places or architectural structures along a familiar path through a familiar environment [see Ross & Lawrence (1968), Yates (1966)]. Each locational image then serves functionally like a peg in the pegword mnemonic system; that is, to learn a list of new items, S images the referent of the *nth* word doing something in the *nth* imaginal location from his list of loci.

Comparing the pegword and loci methods, two apparent differences are (a) the use of static locations vs. objects as pegs, and (b) information about an item's numerical position within the list for the pegword method vs. only serial-order information available in the loci method. But if numerical or order information is not critical to the recall measure, then the sole difference between the methods appears to be whether the pegs are familiar locations as opposed to names of any concrete objects. Assuming that the *S* freely constructs an imaginal relation between the *nth* to-be-remembered (TBR) referent and the *nth* peg, a locational peg may restrict the class of available relations more than does a nonlocative peg. For example, locations cannot be animate actors or recipients of actions, they are typically "large and unmanipulable," and so forth. To assess such matters, a direct comparison of the two methods was made in the following experiment.

A second, subsidiary question concerned whether recall with the mnemonic methods could be degraded by use of especially contrived classes of items. One attempt used the pegwords themselves as occasional list items to be recalled. For example, *shoe* was the second pegword; in, say, the third list to be learned, *shoe* might be the eighth list word (eighth *plate* as a pegword). If *S* forms some image integrating *shoe* with *plate*, he might later recall *plate* incorrectly as the item in the second position. Further, the *plate-shoe* pairing might cause unlearning of the second item previously paired with *shoe* used as a pegword. Clearly, these are the sorts of interpair interference possibilities created by a "double-function" item, acting both as a cue (albeit implicit as is *shoe* here) for one response and as a response to another cue. For such reasons, one might expect recall of those items which happen also to be pegwords to be poorer than control items.

A third question regarding item type involved what might be called the "expansiveness" of the referent of the item to be learned. The notion is intuitive and not sharply defined; but we consider expansive terms to be those whose referents are large, expansive areas without definite points of fixation or focus. Examples are *prairie, ocean, valley, earth, and street*. Intuitively, it would seem to be difficult to manipulate these very much in imagery; most likely, they are used only as background settings in which other actions
occur. One might expect such expansive terms to create particular difficulties for Ss using the method of loci. Although one can easily imagine a small manipulable object placed in a particular locational scene, (e.g., an alligator under the small portico of a campus walkway), it would seem difficult to imagine a large locational expanse (e.g., prairie) at another, smaller location. In fact, it is difficult to envisage how Ss could handle such images to effect learning. In any event, learning of such expansive terms was compared to that of the names of smaller, manipulable objects in the following experiment. For similar reasons, a set of simple “location” names were also used as special items. These locations were not necessarily large expansive areas (e.g., saloon, camp, jail, bungalow). Possibly any locations will be difficult to image and learn by the pegword systems.

**Method**

**Design**

Mnemonic strategy was a between- S variable. One group, the SI (Separate Images) Ss, learned the list of 20 rhyming pegwords (see Table 1) and then learned five successive lists by forming separate peg-to-item images for each list A second group, PE, learned the five lists by progressively enlarging and elaborating the scene imagined around each pegword. A third group, L (loct), used the progressive-elaboration method but with a self-made list of 20 familiar locations as mnemonic pegs. After an initial practice and instructional session, all Ss received one presentation of five 20-item lists, being tested for recall immediately after each list, for recall of all five lists at the end of the learning session, and for retention at a one-week interval. The type of item—whether a location, an expansive term, a pegword, or control word—was a within-list, within-S variable.

**Procedure**

The Ss received three separate 40-min sessions, the first two in groups, the last in an individual meeting with E. The initial session instructed and practiced S on the use of mental imagery for associative learning. In this session, S studied then recalled two lists of 20 noun–noun pairs bearing an A–B, A–C relation to one another. Each S was further trained on his specific mnemonic technique and told how he was to use it for learning multiple lists. Thus, Ss in the pegword conditions were practiced on pegwords rhyming with the first 20 integers (to a learning criterion), and shown how to use these pegwords for learning and recalling new lists.

Regarding learning of successive lists, Ss in the SI condition were told to construct an entirely new mental picture for each successive TBR word paired with a given pegword. For example, to form the pair DOG–BICYCLE, it was suggested that they might picture a large St. Bernard pedaling a small bicycle, for the second pairing, DOG–TOP HAT, to picture a small chihuahua peaking out from under a top hat, for the third pairing, DOG–CIGAR, to picture a grumpy bulldog chomping on a cigar. That is, it was suggested that the imaginal instantiation of the general concept should change as the general term was paired with successive TBR words.

In contrast, Ss in the PE condition were instructed to progressively build up scenes encoding the peg with all the prior TBR terms. They were to recall their prior scene from the pegword, then add the new item into that composite image. Thus, for the examples above,

<table>
<thead>
<tr>
<th>TABLE 1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MNEMONIC NUMBER RHYMES FOR 1–20</strong></td>
</tr>
<tr>
<td>One is a gun</td>
</tr>
<tr>
<td>Two is a shoe</td>
</tr>
<tr>
<td>Three is a tree</td>
</tr>
<tr>
<td>Four is a door</td>
</tr>
<tr>
<td>Five is knives</td>
</tr>
<tr>
<td>Six is sticks</td>
</tr>
<tr>
<td>Seven is oven</td>
</tr>
<tr>
<td>Eight is plate</td>
</tr>
<tr>
<td>Nine is wine</td>
</tr>
<tr>
<td>Ten is hen</td>
</tr>
</tbody>
</table>
after input of the third pair $S$ might have constructed the composite image of a St. Bernard wearing a top hat and riding a bicycle while smoking a cigar.

Subjects in the third condition (L), using the method of loci, made up their own list of locations; it was suggested that they use distinct locations suggested by their route during a typical Wednesday at college, from awakening in the morning in their room through locations strwn along the routine path of their day. The $S$ practiced recalling this list of locations several times. These loci $S$s were also instructed to use the progressive elaboration method for learning successive lists, calling to mind the composite scene at the $n$th location and then adding the new $n$th TBR word to that composite scene.

The $S$s were dismissed from this initial session with instructions to rehearse their pegword (or locations) list several times throughout the day (but not to use it for learning any material), and to return to the following day. The following day, $S$ first recalled and restudied his pegwords or locations, was reminded of how to use these for learning successive lists, and then commenced study-then-test cycles on five successive lists of 20 nouns. Each list was read once by $E$ at a rate of 10 sec per item. Immediately after each list, $S$ attempted to recall the 20 words of that most-recent list, writing his responses on a sheet with blank spaces numbered 1 to 20. The $S$s were asked to recall each item in its right position if they were able; but they were encouraged to guess at position if they recalled a list word but were unsure of its position within the list. Four minutes were allowed for this immediate recall of each list.

Between successive lists, $S$s were reminded of the strategy they were supposed to use to learn successive lists. Also between lists, $S$ was permitted to look back at his list of pegwords or locations if he was unsure of one or another item or if (This occurred only very rarely.) After the immediate-recall test for the fifth list, $S$s were asked to attempt recall of all five lists (100 words), writing each on a large sheet marked off with five columns (lists) of 20 numbered blanks. They were asked to try to recall the list and position of each item recalled, but to guess at these if they could only remember the word. Ten minutes were allowed for this end-of-session recall.

After finishing their recall, $S$s signed up for an individual session one week later and were asked not to think about the materials they had learned, and not to use their mnemonic in the interrim. When $S$ returned a week later, he first recalled his pegword or location list, being prompted and corrected if needed, and then attempted written recall again of the five previously learned lists. Although $S$ was requested to try recalling each item in its correct list and position, he was again encouraged to guess these when he was unsure.

Although $S$s had unlimited time for this recall, all finished within 10 min. Following debriefing, $S$s were paid and then dismissed.

Subjects and Materials

The $S$s were Stanford undergraduates recruited from the Introductory Psychology class. They were paid $5.25 but only if they attended all three sessions as scheduled. The vicissitudes of show-ups and drop-outs left 10 $S$s in the SI group, 12 in the group doing PE with pegwords, and eight in the group using the method of loci with PE.

The word lists were nouns, with ratings of 6.00 or greater on both imagery and concreteness in the norms of Pavio, Yuille, & Madigan (1968). The 20 items in each list were divided into categories, as follows: one pegword, 5 or 6 expansive or locational nouns (these are not mutually exclusive categories), and 13 or 14 control items. The paired-associated practice lists learned in the initial training session were similarly constructed. The rhyming pegwords were all concrete, "nonexpansive" nouns.

Results

Pretraining

Since all $S$s studied two paired-associate lists during pretraining, before differential mnemonic instructions, their performance on those lists provides some indication of comparable learning ability in the three groups of $S$s. Immediate recall of these two lists was practically identical for the three groups, $F(2, 27) = .06$, so the groups may be assumed to be equivalent before differential instructions.

List-Recall Measures

Recall of the 20 word lists may be scored "strictly" or "leniently" according to whether or not a recalled word is counted correct if it is recalled but not in its proper position within the list. In immediate recall, practically all recall was correct by position, so the two scores are nearly identical there. But at both delayed recall tests, the two measures differ, reflecting some intralist and interlist confusion. Despite differences in absolute level between these measures, they both reveal the same differences of interest among experimental conditions. Therefore, our main analyses here will concentrate on "strict" scoring of performance.
Immediate Recall

Recall performance immediately after presentation of each list was uniformly high for all conditions and lists. The Ss' recall in conditions L, PE, and SI averaged 88%, 84%, and 87% correct, respectively. These averages did not differ significantly. Nor was there significant variation in immediate recall across the five lists within the day. Mean percent recall, averaged across conditions, was 86, 84, 89, 89, and 84 for the five lists. So one may conclude that Ss using the different mnemonic strategies are about equally good at unpaced immediate recall, whether ordered or unordered recall is considered.

End-of-Session Test

Differences among the conditions appeared on the MMFR test at the end of the learning session. Mean percentages recalled (strict scoring) for each list and each experimental condition are shown in the left-hand panel of Figure 1. A variety of important (and statistically significant) facts can be read off the set of curves in this left-hand panel. First, of course, there was significant forgetting for all groups, comparing overall performance to the mid-80's levels prevailing at immediate recall. Second, there was more forgetting by Ss in the SI condition than by Ss in the other two conditions ($p < .01$). Performance levels were equivalent for groups PE and L. Third, recall increased across the five lists for the SI Ss (a simple "recency" effect) but not for the other Ss. A linear component for the recall trend across lists differed significantly from zero for the SI group, $t(9) = 4.01, p < .01$, but not for the PE and L groups. To summarize these end-of-session results, then, the SI Ss forgot more than Ss building composite images; the SI Ss also showed strong "recency" across lists in their recall curves, whereas these differences were absent for the PE and L groups.

One-Week Test

The mean percentages recalled (strict scoring) for the various conditions at the one-week retention test are shown in the right-hand panel of Figure 1. A variety of further facts can be read off the pattern of curves in that right-hand panel. First, forgetting has progressed beyond the end-of-session level of the week before, with even more forgetting for group SI than for groups PE and L. Average percentage loss scores over the one-week interval, calculated for items recalled at the end-of-session test, were 69%, 25%, and 24% for conditions SI, PE, and L, respectively. Thus, the conditions

![Figure 1](image-url)

**Fig. 1.** Mean percentages recalled across lists for the three groups during the end-of-session test (left panel) and the one-week delay test (right panel).
differ not only in end-of-session recall but also in differential forgetting of items recalled earlier.

Second, for each condition recall now decreases significantly across the five lists within the training series (a "primacy" effect). Statistical tests compared the linear-trend scores for each group to zero by $t$ tests; significance levels for rejection of the null hypothesis were .05, .001, and .005 for groups SI, PE, and L, respectively. The decline across lists was greater in absolute terms (but less in relative percentage terms) for groups PE and L than for group SI.

Recall by Item Type

Each list contained several special items (viz., pegwords, locations, expansive words) selected in anticipation that Ss using imagery mnemonics might experience unique difficulties in learning such items. Correct ordered recall percentage of each word type—whether location, expansive term, or pegword—was compared to recall percentages of "all other" items in a series of chi-square tests, for the three conditions for the three tests (immediate, end-of-session, one-week delay). Of 27 such comparisons, not a single chi square was statistically significant. In other words, on all measures for all groups, these "special" items were recalled at about the same level as the control items. So our expectations in this regard proved to be entirely misguided.

Types of Errors

To obtain a more complete profile of the performance differences among the mnemonic strategies, the frequency of seven different categories of errors were examined. In immediate recall, omissions, semantic (extralist) intrusions, and within-list intrusions (order or position) errors accounted for, respectively, 53%, 26%, and 13% of all errors, and the three groups did not differ significantly in these percentages $\chi^2(12) = 18.9$, .05 < $p$ < .10. Although omissions increased in absolute level at the delayed tests, the relative percentages of errors that were omissions remained between 50% and 60%. On the delayed recall tests, percentages of errors that were interlist intrusions (i.e., recalled on the correct peg but assigned to the wrong list) increased dramatically, from only 1% on immediate tests to 26% at the end-of-session test and at the one-week. Clearly, "list differentiation" information is lost between the immediate and delayed tests. Although the SI group was slightly favored in their percentage of list-differentiation errors to all errors, this particular difference (of 24% for SI vs. 28% for other Ss) was not statistically significant. There were significant differences among the three conditions in their overall distribution of errors, at both the end-of-session test, $\chi^2(12) = 70.4$, and the one-week test, $\chi^2(12) = 115.3$. Generally, Ss in groups PE and L made relatively more overt errors including semantic (extralist) intrusions, within-list (order) errors, and between-list (differentiation) errors. In scoring leniently and for "gist" rather than literal recall, such recalls would be counted as correct, giving conditions PE and L an even greater advantage at delayed recall relative to condition SI.

 Discussion

These results provide several conclusions. First, the method of loci and the pegword mnemonic are virtually equivalent for the sorts of recall tasks used here. Mental snapshots of familiar locations are just as effective for retrieval cues as are the smaller concrete objects typical of the pegword mnemonic. Though not obvious in advance, the equivalence of the two cue lists is perhaps understandable in retrospect. If one simply asks what characterizes the loci self-selected by our Ss, the reply is that a "locus" was typically a known object or collection of objects, such as a S's bed, the queue of people at the breakfast cafeteria, an artistic water-spray fountain, and so forth. If location images are simply "mental snapshots" of familiar objects, then perhaps we should never have expected any difference
in recall between loci and object pegs such as gun, shoe, tree, and gate. The only apparent difference remaining in the two mnemonics is that numerical position of an item in the list is immediately available with the number-rhyme pegwords whereas only serial-order information is directly accessible for items with the method of loci (i.e., numerical position of an item has to be calculated).

A second conclusion from our results is that instructed learning strategies can radically alter the effects on recall of requiring Ss to recall successive lists that conform to conventional A–B, A–C paradigms of negative transfer. This might be labeled appropriately as "cognitive control of RI." The difference in recall between Ss using progressive elaboration and those constructing independent images for each list was consistently large and dramatic on delayed tests (though not on immediate tests). Although Ss in condition SI were asked to encode the pegword by a different imaginal instantiation each time, this did not help later MMFR for the several lists, as would be implied by Martin's (1968) earlier theory of encoding variability (which assumed that A–C will not cause unlearning of A–B if A is encoded differently in the two instances). The SI Ss suffered large amounts of RI, and showed a typical "RI" function across lists at the end-of-session recall. In contrast, the progressive elaboration strategy induces S to perform frequent recalls (to the pegword) of items on earlier lists, thus effectively chaining together successive items (A–B–C–D···), and repeatedly remembering and rehearsing that expanding chain. The method obviously "keeps alive" in active memory elements from earlier lists. The procedure resembles a discriminated version of Mandler and Dean's (1969) "1 + 1" procedure, whereby a list is learned by giving S one new word per trial, after which he recalls all items up to and including this most recent one.

The notion that progressive elaboration involves retrieval of, addition to, and rehearsal of a serial chain receives several sources of support. First, the "primacy" effect in recall across lists (cf. right panel of Figure 1) for groups PE and L is consistent with differential implicit retrieval and rehearsal favoring the earlier lists. Using the progressive elaboration strategy, by the time the fifth list has been presented, the number of implicit rehearsals of the successive items (to the peg) is roughly 5, 4, 3, 2, and 1 for items in lists 1 to 5, respectively. If resistance to long-term forgetting is a function of the number of prior retrievals and rehearsals (implicit or otherwise), as suggested by Allen, Mahler, and Estes (1969), then this differential rehearsal notion can explain the "primacy" effect at the one-week test for the PE and L groups.

A second line of evidence supporting the "rehearsal of a serial chain" hypothesis is that Ss using this method nonetheless maintained fairly accurate list identification of the successive items. At the end-of-session MMFR test, when Ss in groups L and PE recalled an item on the correct peg, with probability .92 it was assigned to the correct list. Even at the one-week retention test, this conditional probability was still .78, considerably above the chance level of .20. Rehearsal of a serial chain—first peg-to-A, then peg-to-A-then-B, etc.—could account for long-term maintenance of this serial-order information within the same day as presentation. On the other hand, such order information would not be available if S recalled at a later time a single "still-photographic" image in which all objects were represented simultaneously. Clearly, an image of a simultaneous collection of objects, if it is considered something like a still photograph, does not provide recency information from the temporal placing objects into the imaginary scene.

A third finding of this study is essentially negative; we were unable to show any difference in learning between "locational or expansive" terms and control, "small-object" terms (note that, by selection, these had equal imagery ratings in the Paivio et al. norms). This result is also understandable in retro-
spect. Our Ss seemed to experience no difficulty imagining scenes about expansive terms. Typically, they were objects of perception rather than agents or objects of action scenarios, for instance, “from my porch [a locus], I see a deserted prairie stretching into the distance with a tall mountain rising up on the horizon.” So such items are easily imaged, though they seem more to be “still life” than action images. Earlier research by Rohwer (1970) is relevant to the issue: his Ss learned simple subject–verb–object sentences just as rapidly when the sentence contained a stative verb (liked, wanted, feared, etc.) as when it contained an action verb (hit, ran, kissed, etc.). From such considerations, one should not expect locational or expansive items to be inferior in learning to control items.

Finally, there are two puzzles in these data which require brief mention. The first is that items which were pegwords appeared to cause no particular difficulty in recall, despite firm theoretical grounds (re: “double function” lists) for expecting an effect. However, since there was only one pegword item per list of 20, this may have been too small a dose of “double functioning” to create appreciable interference. The issue deserves a more complete experiment since an effect doubtless can be shown in extreme cases, e.g., where the entire TBR list is a reordering of the 20 pegwords.

The second puzzle is the decline in recall across lists for the SI Ss at the one-week retention test (the bottom curve in the right-hand panel of Figure 1). Why should there be this decline across lists for SI Ss, who presumably rehearse each list only at input? Comparing the two curves of the SI Ss, for end-of-session vs. one-week recall, the pattern is reminiscent (except for different time scales) of “negative recency” phenomena reported recently by Craik (1970), McCabe & Madigan (1971), and others. However, extant explanations of the “negative-recency” phenomenon (e.g., immediate “dumping” after recall of items in short-term memory) seem inapplicable to the conditions of the present experiment.

An important consideration is that this decline in recall across lists for the SI group was obtained only with strict scoring (this was not the case for other groups); there was no hint of this decline with lenient scoring, suggesting that the decline was mainly due to poor list identification of items recalled from later lists. The data on this matter are insufficient to guide speculations in a profitable manner.

REFERENCES


(Received January 12, 1972)