Journal of Experimental Psychology:
Learning, Memory, and Cognition
1984, Vol. 10, No. 3, 369–385

Maintenance Rehearsal: A Two-Component Analysis

Moshe Naveh–Benjamin
Ben-Gurion University of the Negev, Israel

John Jonides
University of Michigan

Maintenance rehearsal of verbal material has traditionally been viewed as a unitary process involving the recirculation of information through memory. According to one view, this recirculation results in long-term memory traces; according to another, it does not. We propose an alternative model that views maintenance rehearsal as a two-stage activity. The first stage, which demands effort, consists of various processes involved in retrieving an articulatory program of the to-be-rehearsed material, loading it into a rehearsal buffer, and beginning to execute it. The second stage is hypothesized to be more automatic. It involves repetitive execution of the rehearsal program. According to this model, it is especially the first stage that causes the creation of long-term memory traces. We tested the model by applying three criteria to establish the distinction between a controlled and a more automatic component of rehearsal. The experiments that included these tests also compared maintenance and elaborative rehearsal conditions both on the criteria for distinguishing controlled and automatic processing and on their respective effects on the establishment of long-term memory traces. Overall, the results of these experiments provide support for a two-stage model of maintenance rehearsal and for the assertion that the early stage causes rehearsed material to be stored with some permanence.

During the past 10 years or so, a controversy has developed concerning the effect of rehearsal on the storage of information for long periods of time. There is general agreement that various elaborative rehearsal strategies such as chunking and creating mental images have beneficial long-term memory consequences. The controversy concerns maintenance rehearsal, the strategy of repeating material over and over again.

At one level, the controversy is rooted in theory. Atkinson and Shiffrin (1968) and others have asserted that maintenance rehearsal results in a transfer of information from short- to long-term memory. However, Craik and Lockhart's (1972) levels-of-processing view sees maintenance rehearsal as an operation that holds a particular code intact without creating deeper, more lasting codes.

At another level, there is substantial empirical disagreement about the long-term memory consequences of maintenance rehearsal. A number of empirical studies have found that this strategy results in memory traces that far outlast the rehearsal process iteself (e.g., Dark & Loftus, 1976; Darley & Glass, 1975; Glenberg, Smith, & Green, 1977; Maki & Schuler, 1980; Mechanic, 1962; Modigliani, 1976; Rundus, 1971; Woodward, Bjork, & Jongeward, 1973). However, other studies have found no effect at all of maintenance rehearsal on later retrieval (e.g., Craik, Gardiner, & Watkins, 1970; Craik & Watkins, 1973; Gotz & Jacoby, 1974; Jacoby & Bartz, 1972; Meunier, Ritz, & Meunier, 1972; Modigliani & Seamon, 1974; Roenker, 1974; Rundus, 1977; Tulving, 1966).

Why the empirical discrepancy? Nelson (1977) cited three methodological issues that are relevant to the controversy: lack of statistical power in various studies, lack of control over the rehearsal activity itself, and lack of control over relevant prior memory tests that may affect later retrieval performance.

This list can be supplemented with other

This research was supported in part by National Science Foundation Grant BNS-80-24512, in part by National Science Foundation Grant BNS-79-13757, and in part by Air Force Office of Scientific Research Grant 82-0297.

We thank K. Holyoak, G. Logan, J. Nairne, G. Olson, and K. Watkin for their helpful comments on an earlier draft of this paper. We would also like to thank Dan Perlman, Pat Wilson, and Richard Ruhrold for their help in the data collection and analysis.

Requests for reprints should be sent to Moshe Naveh–Benjamin, Department of Behavioral Sciences, Ben-Gurion University of the Negev, P.O. Box 653, Beer–Sheva, 84120 Israel.

problems as well. First, there are large differences among the various tasks that have been used to engage maintenance rehearsal. One may well imagine that some tasks are more successful than others in having subjects just repeat the to-be-tested material without trying to retain it. Second, some experiments have used recall as a measure of later retention whereas others have used recognition, tests that may differ greatly in sensitivity to stored traces of the rehearsed material. In fact, Woodward et al., (1973) and Glenberg et al. (1977) demonstrated effects of maintenance rehearsal on later recognition but not on recall.

Then there is the problem of how to define maintenance rehearsal operationally. It will not do, of course, to define it by its lack of consequence for long-term memory because this is precisely what is at issue. One feature that is clearly traditional in a definition of maintenance rehearsal is repetition. Operationally, though, this is insufficient, because it is difficult to observe whether subjects are engaged merely in repetition of a body of material. We can supplement this feature with the requirement that maintenance rehearsal be engaged exclusively for the short-term manipulation of information (Nairne, 1983). Finally, we and others have added to the repetition criterion the feature that maintenance rehearsal not place heavy demand on central processing capacity (Glenberg & Adams, 1978; Naveh-Benjamin & Jonides, in press). However, this too is problematic. As we have argued, Glenberg and Adams (1978) might have called on too stringent a criterion of light capacity demand (Naveh-Benjamin & Jonides, in press). In the absence of some absolute criterion for what a heavy or a light demand is, we prefer to define the presumed light demand by comparison with elaborative rehearsal, which is assumed to place a heavier demand on resources. Later, we assert that even within the act of maintenance rehearsal, capacity demand varies; nonetheless, the act as a whole, by definition, is asserted to require a lighter capacity usage than elaborative rehearsal.

Defined in this way, maintenance rehearsal does seem to establish traces with some durability (Glenberg & Adams, 1978; Naveh-Benjamin and Jonides, in press). However, the available evidence suggests that it is the early segment of a maintenance rehearsal act that produces long-term memory effects. This was shown in our earlier work in which the first few rehearsals in a sequence were shown to have an effect on later recognition performance for the rehearsed material, whereas later rehearsals had a relatively small effect (Naveh-Benjamin & Jonides, in press). Also, in a recent study, Nairne (1983) showed that associative learning during rote rehearsal is limited to the initial stage of the rehearsal interval. These patterns of results lead us to propose a two-stage model of maintenance rehearsal, only one stage of which has a significant effect on the creation of long-term memory traces. The first stage occurs early in the rehearsal sequence. It consists of various processes that are required to retrieve an articulatory code of the to-be-rehearsed material, to program this code for execution, to set up timing for recirculation of the code, and to fine tune its first few executions.

The second stage is a product of the first. After an item has been rehearsed several times, we presume that the act of rehearsing it becomes more automated. Now, some may object that the literature concerned with the automaticity of cognitive processes indicates that it takes extensive practice to automate a mental activity (e.g., Schneider & Shiffrin, 1977), so how could rehearsal become automated in a matter of seconds? A plausible response to this is that highly practiced skills such as speech production are much more susceptible to automation than are newly learned skills when the parameters of the produced speech are held constant (see Logan, 1979, and MacKay, 1982, for discussions of the speed with which some aspect of processing may become automated).

What are the implications of this model of maintenance rehearsal for tests of memory for the rehearsed items? A recent line of research lends support to the hypothesis that the amount of cognitive effort or attention that is devoted to a task has a direct effect on memory performance for items in the task (e.g., Eysenck & Eysenck, 1979; Johnston & Heinz, 1978; Tyler, Hertel, McCallum, & Ellis, 1979). This finding is made especially relevant to maintenance rehearsal by the finding of Geiselman, Woodward, and Beatty (1982), who found that cognitive effort helped maintenance rehearsers build long-term memory traces.

Suppose that the extent to which relatively permanent traces are laid down by maintenance rehearsal is a function of the degree to which this rehearsal is demanding of cognitive capacity. Because, by hypothesis, the first stage of rehearsal demands more cognitive capacity than the second, we would expect to find a larger effect of the first stage on later memory performance than of the second stage. This pattern of results has been obtained in our previous research (Naveh–Benjamin & Jonides, in press), in the work of Geiselman and Bjork (1980), and in research by Nairne (1983).

The present series of experiments further tested this model in two ways. The first test concerned the later memory consequences of rehearsal. The model predicts that for maintenance rehearsal, an increase in the number of repetitions of an item early during the rehearsal period would result in the creation of a more accessible memory trace for later retrieval. However, further rehearsals would have little if any further effect on strengthening the trace. To evaluate this prediction, we compared the effects of maintenance and elaborative rehearsal, because by hypothesis, the continued effort devoted to elaborative rehearsal throughout the rehearsal period should have a progressively greater effect on the retrievability of the rehearsed item.

The second test of the model involved a comparison of early and late stages of maintenance rehearsal (repetition) using a series of criteria to distinguish whether rehearsal is automatic. In order to provide a baseline for this comparison, we included an elaborative rehearsal condition in which subjects were required to rehearse the memory items effortfully throughout the rehearsal period. The model predicts that the automaticity criteria should discriminate between the two hypothesized phases of maintenance rehearsal, compared with a baseline provided by the elaborative rehearsal condition.

In order to execute the tests concerned with automaticity, we applied three criteria that have been used to test whether a process is automatic (see Jonides, 1981, and Posner & Synder, 1975, for discussions of these criteria). Perhaps the most frequently cited criterion for an automatic process is its relatively light demand for cognitive resources. In the present context, this can be interpreted to mean that Stage 1 of maintenance rehearsal demands more processing resources than does Stage 2. To test this hypothesis, we used a multiple-task procedure in which subjects were responsible for detecting a probe stimulus during the intervals when they were rehearsing (see Stanners, Meunier, & Headley, 1969, for a similar procedure). Examining probe response time (RT) as a secondary task at various points in the rehearsal sequence enabled us to assess changes in the capacity demands of maintenance rehearsal (compared with elaborative rehearsal) throughout its execution.

Our second criterion for automaticity concerns the extent to which rehearsals become stereotyped as they continue. Typically, as skills or routines become progressively more automatic, they become more uniform and unvarying in their characteristics. In fact, it is precisely this decrease in variance that is one of the key benefits of automation: It permits predictable execution of a skill, thus allowing the performer to dispense with higher level control (a fact for which professional golfers among other athletes are surely thankful). In the first experiment, we examined the development of stereotypy in each rehearsal sequence by measuring variability in the duration of the rehearsal utterances. Our prediction was that variability would decrease with more rehearsals, but only for maintenance rehearsal, not for elaborative rehearsal.

The third criterion for automaticity has been recognized at least since the classic experiments of Stroop (1935): In general, subjects have a difficult time interrupting an automatic process once it has begun. For maintenance rehearsal, we translated this criterion into a prediction that as overt utterances of a rehearsal sequence continued, subjects would find it progressively more difficult to interrupt these utterances in order to produce a different utterance. Once again, we tested this prediction by comparing maintenance rehearsal with elaborative rehearsal.

Overall, then, the experiments had two major goals: to examine the memory consequences of the two hypothesized stages of maintenance rehearsal and to test whether early and late portions of a maintenance rehearsal sequence could be discriminated according to their degree of automaticity.

Experiment 1: Processing Capacity and Stereotypy

In order to meet the research objectives outlined above, we adopted a paradigm developed by Glenberg and Adams (1978) that we had used previously (Naveh–Benjamin and Jonides, in press). In this paradigm, subjects' main task on each trial was to hold 3 two-digit numbers in memory briefly and then serially recall them, as in the Brown–Peterson task. As a distractor task during each retention interval, subjects were presented a pair of words to repeat over and over again, with different word pairs used for different intervals. After all trials were completed, subjects were required to respond to a forced-choice, cued recognition test for the words that had been recited during the retention intervals.

This paradigm corrects for various methodological problems of previous studies: First, in order to elevate performance above the floor, recognition, rather than recall, was employed as an assessment of later retrieval. Second, a sufficient number of subjects and trials were included to provide power to detect the somewhat larger effects that were expected with a recognition test. Third, subjects were required to repeat the words aloud, so that their rehearsal activity could be monitored. Fourth, we used a demanding primary task (storing 3 two-digit numbers) to keep attention diverted from rehearsal. Finally, no initial assessment of retrieval was given for the rehearsed words, so that the final recognition test was uncontaminated by previous tests.

Note that if subjects are not told of the later recognition test, the paradigm provides a good opportunity to test for incidental learning of words that have been stored, using only the maintenance rehearsal strategy of repetition. This feature of the technique was further enhanced by subjects' belief that recall of the numbers on each trial was their principal task. So it was to their advantage, when they did not expect the later recognition test, to forget the words after each rehearsal period ended. This incidental learning-maintenance rehearsal condition was contrasted with another condition, in which subjects were told about the later recognition test and were instructed to use an elaborative rehearsal strategy for remembering the words.

The present experiment tested two criteria to discriminate effortful from automatic processing during maintenance rehearsal: the extent of draw on processing resources and the degree to which rehearsal utterances become stereotyped. We tested the first criterion by examining RTs to probes that appeared while subjects were rehearsing. For maintenance rehearsal, we predicted that early during a sequence, these RTs would be high relative to the RTs late in a sequence. For elaborative rehearsal, we predicted no systematic change. The second criterion was tested by examining the variability of rehearsal utterances early and late in a sequence. Our model predicted decreasing variability for maintenance rehearsal and no change for elaborative rehearsal.

Method

Design. The experiment included two major independent variables. The first, varied within subjects, was the number of rehearsals required on each trial (1, 4, or 10). The second, varied between subjects, was the instruction under which rehearsal occurred. In one condition, subjects were told simply to repeat the words (maintenance rehearsal, with the later recognition test being a test of incidental memory); in the other, they were told to create images linking the words and to elaborate these images to prepare for a later memory test (elaborative rehearsal, with intentional memory for the rehearsed words).

There were several dependent variables in the experiment. We measured recall of the numbers, characteristics of the overt rehearsal utterances of the words, probe RT performance, and later recognition performance.

Subjects. Fifty students at the University of Michigan served as paid volunteers in a session approximately 90 min long. The subjects were each paid $5 for their participation.

Apparatus. Stimuli were displayed on the face of a VT-52 terminal that was interfaced with a PDP-11/60 computer. The terminal was located in a lighted, soundproof booth. A small loudspeaker was placed in front of subjects. A portable Sony tape recorder with a condenser microphone was used to record the sessions. The stimuli for the cued recognition test were typed on five sheets, which subjects completed by hand.

Stimuli. There were four types of stimuli that were employed in each experimental session:

1. Numbers: From the subjects' point of view, the main experimental task was to memorize and recall sets of 3 two-digit numbers. These numbers were selected randomly with the following constraints: All stimuli were drawn from the numbers 12–98 inclusive; no numbers whose two digits were identical (e.g., 55) were included; no two numbers in a set were chosen from the same decade; no two numbers in a set were integer multiples of one another. With these constraints, the numbers in each set were randomly selected for each subject.

2. Words: Throughout each interval during which the numbers were to be retained, subjects were required to

recite aloud (rehearse) a pair of words. The words chosen for rehearsal were all one-syllable words of three to five letters in length and of greater frequency than 20 occurrences per million (Kučera & Frances, 1967). In addition, they were paired so that the words in a pair neither rhymed nor had any obvious semantic relationship. Otherwise the selection and pairing of words was accomplished randomly.

3. Probes: On 66% of the trials, a visual probe (an asterisk in the center of the display) appeared during the retention interval on the face of the terminal. Subjects were required to respond to this probe as fast as possible while continuing rehearsal by pressing the space bar on the keyboard of the terminal.

4. Recognition test: After all experimental trials were completed, subjects were given a four-alternative, cued recognition test for the words that had been recited during the rehearsal intervals. Each recognition trial consisted of a cue word that was randomly chosen from the two words that had been rehearsed in each rehearsal pair. This word appeared on the left margin of the page. On the right, there were four words. One of these was the other word in the pair and the others were distractor words that had been chosen using the same principles used to select the words for rehearsal. These distractor words had not been used in the first phase of the experiment. The order of the four words for each recognition trial was randomized with respect to position on the answer sheet.

Trial structure. There were 130 trials in the experiment. Of these, the first 13 were considered practice and were not analyzed. Furthermore, the next 6 trials after the practice trials and the last 5 trials of the experiment were also excluded from analysis in order to minimize the effects of primacy and recency in the later recognition test. Thus, there remained 106 test trials. On each of these trials, and on the remaining 24 that were not included in the analysis, the following sequence of events occurred:

First the word *ready* appeared for 2 s, centered on the face of the terminal. Following this, 3 two-digit numbers appeared for 4 s; they were horizontally arrayed and centered on the screen. The numbers were then replaced by two words, also centered and horizontally displayed, that remained in view for 0.6 s.

Following presentation of the words, subjects heard a set of tones through a loudspeaker in front of them. Each tone was 1000 Hz and was presented for 75 ms. There was a stimulus onset asynchrony of 1,200 ms between successive tones. There were 1, 4, or 10 tones presented. Each length of tone sequence appeared with equal probability, and the various lengths were distributed randomly in a session. This design was intended to prevent subjects from predicting the length of a rehearsal interval. We chose the intermediate length of 4 after results from pilot studies suggested that rehearsing automated after about 4 rehearsals. Following the tones, the word *recall* appeared on the screen. Subjects could then type their recall of the three numbers. After making any necessary corrections, subjects pressed the *return* key on the terminal at their leisure. This initiated a 5-s interval before the next *ready* signal appeared.

During presentation of the tones, a probe might have appeared on the screen. The probe appeared 250 ms[1] after the onset of either the 1st, the 4th, or the 10th tone in the sequence. In all trials, the probe appeared after the last tone in the sequence (i.e., during the last rehearsal).[2]

Table 1
Frequencies of Probes at Various Positions in the Rehearsal Sequence

	No. of rehearsals		
Probe appearance	1	4	10
During 1st rehearsal	39	0	0
During 4th rehearsal	0	26	0
During 10th rehearsal	0	0	13
No probe	0	13	26

The choice of 250 ms as the stimulus onset asynchrony with respect to the tone was based on results of pilot studies that indicated that such a latency period would ensure that the responses to the probes would be made, on the average, roughly simultaneously with the overt rehearsal and hence would fulfill the requirements of a dual task.

The overall probability of probe appearance was 66%. The probabilities of probe appearance in the various rehearsal positions were 33% for a probe in Position 1, 22% for a probe at Position 4, and 11% for a probe at Position 10. With these values for overall probabilities, the conditional probabilities of probes given that subjects knew that they were about to engage in their 1st, 4th, or 10th rehearsals were each 33%. The actual distribution of trials with probes in various positions in the rehearsal sequence is presented in Table 1. This distribution was constructed to prevent subjects from predicting when a probe would be presented.

Procedure. Subjects were run individually. They began a session with instructions about the structure of a trial: They were told that they would be presented with three numbers to hold in memory for a variable period of time, during which they would be required to rehearse a word pair into a microphone at a pace set by the tones. They were told to respond to the tones as quickly as possible without making anticipatory errors, and to respond to the probe when necessary while continuing the rehearsal. In addition, subjects were told about the sequence of events on each trial and about how to enter their number recall on the keyboard.

In the incidental-maintenance rehearsal condition, the task instructions requested subjects to concentrate on the number task, to keep pace with the rehearsal tones, and to devote any spare processing capacity to the probe task.

[1] In pilot studies in our laboratory, we have found no difference in either word recognition or number recall for trials with a probe compared to trials with no probe.

[2] The probe remained in view until the subject responded. If the response time was longer than 750 ms, subjects were presented with a message to this effect, and the trial ended; otherwise, their responses to the probes (depressing a space bar key) were not accompanied by feedback. Responses of greater than 750 ms (these accounted for 5% of all responses) were not included in the analyses reported later. These excluded responses were distributed equally between the two conditions.

Nothing was mentioned about the word recognition task. In the intentional-elaborative rehearsal condition, subjects were told about the later recognition test for the words, and they were urged to rehearse the words elaboratively in order to remember them. Specifically, it was suggested that they create a mental image on each trial such that the images of the words in each pair interact with each other. They were also told that while continuing to rehearse, they should try either to change the interactions between the images or to make the images more vivid. They were told that previous research had shown that such strategies helped subjects in the later recognition test. They were cautioned, however, to keep their recall on the number task at a reasonable level. Finally, as in the other group, they were told to keep pace with the rehearsal tones and to devote any spare processing capacity to the probe task.

Following these instructions, there were 13 practice trials, during which the experimenter corrected the subjects on any errors that were made. Then the subjects proceeded with the remaining 117 trials. The experimenter continued to monitor subjects' performance on the task via an intercom and corrected any deviations from the instructions. All sessions, including auditory stimuli and rehearsal utterances, were recorded on magnetic tape.

After completing the experimental trials, subjects called the experimenter, who then administered the recognition test. Subjects had as much time as they needed to complete this test.

Each session ended with a questionnaire that inquired about subjects' introspections, about their compliance with the instructions (e.g., whether they followed the elaboration instructions in the intentional condition), and also about whether they had anticipated the recognition test.

Results

We begin by presenting data for the number recall, word recognition, and probe RT tasks. Following this, we report the results of several other analyses that bear on the hypotheses. The results reported include only 44 of the 50 subjects. Six subjects were excluded from further analysis. Two of them had recall rates of 0%. The other 4 did not obey the instructions of their respective conditions: 2 in the intentional condition reported in the posttest questionnaires that they did not expect the final recognition test and therefore did not use imagery as an elaborative technique; the other 2 subjects in the incidental condition were excluded because they had suspected the final word recognition test and used elaborative techniques.

Number recall. Table 2 presents the percentage of numbers recalled as a function of rehearsal condition and number of rehearsals. In tabulating these data, a strict criterion was used in which a response was scored as correct only if all 3 two-digit numbers were recalled in their correct order of presentation.

Table 2
Mean Percentage of Correct Responses for Number Recall as a Function of Number of Rehearsals and Rehearsal Condition

Rehearsal condition	No. of rehearsals: 1	*SD*	4	*SD*	10	*SD*
Incidental maintenance	64.7	20.3	52.7	22.1	47.0	27.6
Intentional elaborative	45.8	26.7	32.8	28.4	29.9	28.5

Table 2 shows that performance was at a reasonable level, high above chance (which is only negligibly higher than 0%). Even after 10 rehearsals, subjects on the average correctly recalled the numbers on about 38% of the trials.

The data in Table 2 reveal two trends that were confirmed by an analysis of variance (ANOVA). (An alpha level of $p < .05$ is used for all tests reported in this article.) First, recall performance was significantly higher for the incidental condition, $F(1, 42) = 6.41$. Second, recall performance dropped with increasing number of rehearsals, $F(2, 84) = 31.62$. There was no significant interaction effect of the two variables, $F(2, 84) = 0.21$. Both of the main effects are to be expected in a task of this sort. Because the number task is the primary one for the subjects in the incidental condition, they apparently pay more attention to it in comparison with subjects in the intentional condition, who have to split their attention between this task and the word elaboration task. Also, the standard effect of a distractor activity on number recall is to cause interference with the retained numbers and hence to cause a decrease in number recall performance with extended rehearsal periods (see Naveh-Benjamin and Jonides, in press, for a discussion criticizing the use of no change in number performance as a criterion of maintenance rehearsal).

Word recognition. Table 3 presents the percentage of correct responses on the final recognition test. Inspection of these data reveals that recognition performance was above chance, $t(42) = 8.53$, a result that confirms the claim that recognition is sufficiently sensitive to detect the existence of long-term memory traces in this task.

Table 3
Mean Percentage of Correct Responses for the Word Recognition Test as a Function of Number of Rehearsals and Rehearsal Condition

Rehearsal condition	No. of rehearsals 1	SD	4	SD	10	SD
Incidental maintenance	42.4	8.6	48.0	9.5	50.4	11.5
Intentional elaborative	55.3	9.3	61.5	10.1	69.0	11.9

The proposed two-stage model of maintenance rehearsal predicts that early rehearsals in a sequence, being nonautomatic, cause memory traces of the rehearsed items to be strengthened for later retrieval, whereas the more automatic later rehearsals in a sequence have little if any effect on trace strength. By contrast, elaborative rehearsal is presumed to cause a steady increase in trace strength within reasonable limits. To test these predictions, we examined two planned comparisons of the Rehearsal Type × Rehearsal Interval interaction: one with rehearsal interval 1 versus 4 and one with rehearsal interval 4 versus 10. These two planned comparisons of interactions were conducted with the modified Bonferroni test; they had the following critical value: $F(1, 42) = 4.08$, $p = 0.05$. For details of these analyses, see Keppel (1982).

The Rehearsal Type × Rehearsal Interval interaction (for 1 vs. 4 rehearsals) was not statistically significant, $F(1, 42) = 0.89$, $MS_e = 47.05$. The comparable interaction for the interval of 4 versus 10 rehearsals was reliable, $F(1, 42) = 4.73$, $MS_e = 56.39$. This pattern of results is precisely that predicted by the model. Although the effects are relatively small, they are consistent with pilot data as well as with the results of the second experiment.

It is perhaps not surprising to discover also that, overall, recognition performance was superior in the intentional-elaborative condition than in the incidental-maintenance condition, $t(42) = 5.49$. This effect is especially impressive because it was obtained despite the fact that the task of the intentional-elaborative group involved the distraction of overt rehearsal in addition to the creation of imagery. Overall, the results of our word recognition measure agree quite well with those previously obtained by Meunier, Kestner, Meunier, and Ritz (1974), who used a quite different paradigm to measure long-term effects of rehearsal.

Probe responses. Table 4 presents the probe RT data as a function of rehearsal condition and number of rehearsals. The coefficient of variation for these data (less than 10%) is certainly within the expected range for RTs. This is especially impressive because subjects viewed the probe task as secondary in importance.

As with the recognition data, two planned comparisons were used to test whether probe RT performance conformed to the predictions of the two-stage model. The Rehearsal Type × Rehearsal Interval interaction (for intervals 1 vs. 4) showed no reliable effect, $F(1, 42) = 2.91$, $MS_e = 67.05$. By contrast, there was a reliable effect for the interval 4 versus 10, $F(1, 42) = 8.58$, $MS_e = 79.9$. A planned comparison to assess the overall effect of rehearsal type on probe RT did not reveal a reliable difference, $t(42) = 1.73$. We can explain these results with the assumption that there is a fixed total capacity of processing resources devoted to all three tasks (number, rehearsal, and probe) for both conditions. This result then indicates that subjects must have drawn resources from either the number or the rehearsal tasks in the second stage of the incidental-maintenance condition in order to enhance their performance on the probe task. But we also know, from an analysis of the data in Table 2, that there was no reliable difference in trend for number task performance as a function of number of rehearsals across the two conditions. So we must conclude that the improved probe performance of the incidental-maintenance group between 4 and 10 rehearsals (relative to the elaborative group) came from attentional resources freed from the rehearsal process, which in turn resulted in no

Table 4
Mean Manual Probe Reaction Times (in ms) as a Function of Number of Rehearsals and Rehearsal Condition

Rehearsal condition	No. of rehearsals 1	SD	4	SD	10	SD
Incidental maintenance	507	46	495	41	473	36
Intentional elaborative	504	51	499	47	513	47

further increase in recognition performance during this interval.

These results support the hypothesized model of maintenance rehearsal. Working with the assumptions underlying multiple-task methodology, the results lead one to conclude that there is no detectable change in capacity allocated for rehearsal during the first few maintenance rehearsals, but after maintenance rehearsal becomes more automatic, by hypothesis, there is a decrease in capacity allocated for the rehearsals that is indicated by a faster RT to the secondary probe task. By contrast, in the intentional-elaborative condition in which subjects elaborate on the rehearsed words, there is no change in capacity allocated for rehearsal throughout the interval.

The decrease in probe RT from one to four rehearsals in the incidental condition, although not significant, is a bit worrisome. We expected no change in capacity allocated to the first few rehearsals and therefore no change in RT. This trend may be due to the fact that in trials with one rehearsal, subjects had to switch quickly from inspecting the displayed words to reciting them while still retaining the numbers. This might have inflated the RT for trials in which the probe appeared during the first rehearsal. To minimize this disruptive effect, in Experiment 2 we had probes appear after two rehearsals rather than after one rehearsal.

Trade-off analysis. One may argue that performance in the present experiment merely involves a trade-off between capacity allocated to the number retention task and to the word rehearsal task. According to this argument, changes in performance in word recognition may be due to changes in capacity allocated by the subject to number memorization with increases in the rehearsal interval. Specifically, one may claim that the increase in performance in word recognition from one to four rehearsals is not due to characteristics of the rehearsal process itself, but rather is simply correlated with the decrease in number performance from one to four rehearsals. In other words, it is possible that subjects switch some of their attention after one rehearsal from the number task to the rehearsal task, leading to improved performance on the latter. There are two ways to test this possibility. The first involves correlating (across subjects) number recall with word recognition performance as the rehearsal interval changes. A trade-off hypothesis would predict that these correlations (especially between one and four rehearsals) should be negative, because as digit recall decreases between, say, one and four rehearsals, recognition performance should increase. Table 5 presents correlations that test for a trade-off across subjects. The upper-left value in the table, for example, represents a correlation calculated across subjects. One value entered into this correlation is the difference in number recall between conditions of one and four rehearsals. The other value is the difference in word recognition between conditions of one and four rehearsals. The table shows no consistent trend toward negative correlations be-

Table 5
Correlation Between Difference Scores for Number Recall and Word Recognition Tasks for Each Rehearsal Condition

	No. of rehearsals[a]	
Rehearsal condition	1–4	4–10
Incidental maintenance	−0.06	0.38
Intentional elaborative	−0.12	0.21

[a] From which difference scores were obtained.

There is, however, an even more powerful analysis to test for such a trade-off, an analysis that is based on assessing the relationship between recall and recognition for each number of rehearsals required. Consider, for example, 1 subject's recall and recognition data from all trials in which he or she engaged in 10 rehearsals. If that subject was trading resources between retaining the numbers and rehearsing the words, such a trade-off would be detectable by examining each trial to test whether the numbers were recalled correctly and whether the words were recognized correctly. If there were a trade-off, then there should be relatively many trials on which the numbers were recalled and the words were not recognized, and vice versa. But there should be few trials in which the subject was successful or unsuccessful at both tasks. Thus a ratio of the number of trials of the latter type divided by the number of the former would be less than 1.0. Such an analysis could help us to discover whether a trade-off between resources devoted to the number and rehearsal tasks could account for an increase in later recognition with increased rehearsals. To do this analysis, we must partition the data of the experiment into separate contingency tables for each number of rehearsals required in the task. Then an assessment of the trade-off can be made for each table, and more importantly, one can examine any change in the trade-off as a function of number of rehearsals. This analysis can be performed safely here because the values on the two variables are far from floor or ceiling (when near floor or ceiling, there would be only a small number of observations in one or more cells of the contingency table, thus compromising the representativeness of the analysis). Details of this analysis are discussed by Naveh–Benjamin and Jonides (in press).

Table 6 displays such odds ratios averaged over subjects (averaging was accomplished using log odds ratios then transforming back to odds ratios).

No odds ratio differs significantly from 1, and no ratio differs from the others. This indicates that there were no changes in trade-offs between the two tasks across the different rehearsal intervals.

Analysis of stereotypy. Our next analysis tested whether maintenance rehearsal utterances become more stereotyped as they continue. To test this, we drew a random sample of 20 subjects, 10 from each rehearsal condition. For each subject, 20 trials with 10 rehearsals (without a probe) were randomly selected. These rehearsals were analyzed through a Mingograph machine that gave time–amplitude waveforms of the speech patterns. With these representations of the waveforms, we examined two characteristics of the rehearsal sequences: duration of utterances and variability of utterance durations. For the duration measure, we compared the mean duration of the first 4 rehearsals in a sequence with the mean duration of the last 4 rehearsals. We hypothesized that the last several rehearsals in a sequence would be produced more automatically and be more stereotyped than the first few rehearsals. This may result in shorter, more compact utterances later in a rehearsal sequence.

Table 6
Mean Odds Ratios of Recall Versus Recognition Measures as a Function of Number of Rehearsals and Rehearsal Condition

Rehearsal condition	No. of rehearsals 1	4	10
Incidental maintenance	1.08	0.99	1.15
Intentional elaborative	1.07	0.94	0.95

Table 7
Mean Durations and Mean Standard Deviations (in ms) of the First Four Versus the Last Four Rehearsals in a Given Trial as a Function of Rehearsal Condition

Rehearsal condition	First		Last	
	M	*SD*	*M*	*SD*
Incidental maintenance	558	28	561	19
Intentional elaborative	638	28	631	27

To examine variability, we compared the mean of the standard deviations (of durations) of the first four rehearsals in a sequence with the mean of the standard deviations of the last four rehearsals. This analysis tested the hypothesis that increasing automaticity would lead to stereotypy, in the sense of less variable utterance durations. The intentional-elaborative condition was used as a baseline against which to compare these predictions for the incidental-maintenance condition.

The duration of the rehearsal of a given pair of words was measured from the beginning of the utterance of the first word of a pair to the end of the utterance of the second word in the pair. These durations were analyzed independently by two judges who were not familiar with the hypotheses and who were highly reliable in their judgments. Table 7 presents the mean duration and the mean of the standard deviations of rehearsal durations as a function of rehearsal condition and number of rehearsals.

First, let us examine the duration data. The duration of the rehearsals in the intentional-elaborative condition proved to be longer than that in the incidental-maintenance condition as determined by a two-way ANOVA, $F(1, 18) = 8.23$. This trend may reflect more effort paid to the rehearsals in the elaborative condition. There was, however, no difference between mean duration of the first four and the last four rehearsals, $F(1, 18) = 1.82$, nor was there a reliable interaction effect, $F(1, 18) = 1.71$.

For the measure of duration variability, there was no difference between the two rehearsal conditions, $F(1, 18) = 1.63$. However, there was a significant decrease in variability of durations from the first four to the last four rehearsals, $F(1, 18) = 4.79$. This decrease was predicted for the incidental-maintenance condition due to increasing stereotypy of the rehearsal process after the first few rehearsals. In fact, the significant interaction term from the ANOVA indicated that the decrease in duration variability was larger for the incidental-maintenance condition than for the intentional-elaborative condition, as the table suggests, $F(1, 18) = 4.57$.

Discussion

The overall pattern of results obtained in this experiment provides support for a two-stage model of maintenance rehearsal. In measuring the allocation of processing capacity during rehearsal, we found that probe RT decreased with later rehearsals, as would be expected with increasing automaticity. Furthermore, this decrease in RT was associated with no improvement in recognition performance after an initial improvement due to the first few rehearsals. Also, the decrease in number task performance with more rehearsals proved not to be associated with changes in recognition performance. For the elaborative rehearsal condition we obtained, as expected, no change in RT to a probe with increasing number of rehearsals. The pattern of RT results for this condition was accompanied by a monotonic improvement in recognition performance, presumably due to continued allocation of attention to the rehearsed words.

The second major finding of this experiment concerned the stereotypy of rehearsal. We obtained a larger decrease in variability in utterance duration for maintenance than for elaborative rehearsal, once again suggesting increased automation of the maintenance rehearsal utterances.

Taken together, these results support the claim that maintenance rehearsal is composed of an effortful and an automatic stage whereas elaborative rehearsal is more unified in character. To provide converging evidence for this emerging picture, we tested a third criterion of automaticity in the next experiment.

Experiment 2: Resistance to Interruption

The present experiment tests whether maintenance rehearsal becomes more resistant to interruption as it moves from a hypothesized effortful stage to an automatic stage. The criterion of resistance to suppression or interruption has been used several times previously as a measure of automaticity (Jonides, 1981; Logan, 1981; Palmer, Jonides, & Palmer, 1980; Posner & Synder, 1975). In the present experiment, we tested this criterion by interrupting subjects at various times during the rehearsal interval to test how easily they could stop their rehearsal of the words. According to the proposed model, when subjects are using maintenance rehearsal they should become more resistant to interruption as the rehearsal becomes more automatic. As in Experiment 1, we assessed this prediction by comparing a maintenance rehearsal condition with an elaborative rehearsal condition.

Method

Design. The design was identical to that employed in Experiment 1 except that the levels of the number of rehearsals variable were 2, 4, and 10 rather than 1, 4, and 10, for reasons previously discussed.

Subjects. Fifty-four students at the University of Michigan served as paid volunteers in a session approximately 90 min long. The subjects were each paid $5 for their participation.

Apparatus. The apparatus was the same as that in Experiment 1 except that the display terminal was a Hewlett-Packard Model 1301, which was interfaced with a PDP-11/34.

Stimuli.

1. Numbers and words: The numbers and words were identical to those used in Experiment 1.

2. Probes: Instead of an asterisk, as in Experiment 1, the probe was a visually presented word. Twenty-six 1-syllable, four-letter nouns were selected. They were balanced for word frequency, concreteness, meaningfulness, and imageability with the rehearsed words. Subjects were required to voice these word probes when they appeared. To minimize variability in detecting utterances to the probes, probe words were selected so that they all had the initial voiced consonants *B, D,* or *G.* Five sets of these words were compiled and a set was assigned randomly for each subject, one word per probe trial. In order to operationalize the criterion of resistance to interruption, we

had subjects stop their rehearsal (interrupt it) as soon as they saw the probe word flashed on the face of the terminal and say aloud the probe word as quickly as possible. This was in contrast to the first experiment in which they had to react manually to the probe while continuing rehearsal. The probe word was presented in the center of the screen.

3. Recognition Test: The recognition test was identical to the one used in Experiment 1.

Trial structure. The trial structure was identical to that used in Experiment 1 (including the probabilities of probe appearance) except for the following changes:

1. The number of tones for each trial were 2, 4, or 10, instead of 1, 4, or 10.
2. Pilot results showed that in order to maximize the probability of subjects interrupting themselves to voice the probe while rehearsing the pair of words, the time between presentation of the tone and presentation of the probe had to be reduced from 250 ms to 175 ms.
3. The tones were 100 ms in duration.
4. The probes remained in view for 1,000 ms, whether or not subjects responded.

Results

Results are reported here for 44 of the 54 subjects. The remaining subjects were excluded from further analysis for the following reasons: 2 of them had recall rates of 0%; 3 subjects in the intentional condition reported in the posttest questionnaire that they had not expected the final recognition test; 1 subject in the incidental condition suspected the final word recognition test; finally, 4 subjects were not allowed to complete the experiment because they did not adequately comply with the rehearsal and interruption instructions.

Number recall. Table 8 presents the percentage of numbers recalled as a function of rehearsal condition and number of rehearsals. As in Experiment 1, a strict criterion was used in tabulating these data, so that a response was scored as correct only if all 3 two-digit numbers were recalled in their correct order of presentation.

Table 8
Mean Percentage of Correct Responses for Number Recall as a Function of Number of Rehearsals and Rehearsal Condition

Rehearsal condition	No. of rehearsals 2	SD	4	SD	10	SD
Incidental maintenance	45.9	18.7	41.5	20.8	37.6	23.9
Intentional elaborative	37.3	24.2	31.1	25.6	30.3	27.1

Table 9
Mean Percentage Correct for the Word Recognition Test as a Function of Number of Rehearsals and Rehearsal Condition

Rehearsal condition	No. of rehearsals 2	SD	4	SD	10	SD
Incidental maintenance	45.1	10.4	52.3	11.1	54.1	14.8
Intentional elaborative	57.0	11.7	66.0	12.2	74.2	13.1

The results indicate recall performance was lower than comparable performance in Experiment 1. This may be due to the probe task in this experiment interfering with the recall task more than did the probe task of Experiment 1; in Experiment 2, subjects had to interrupt their rehearsals and utter the verbal probe, whereas they had only to respond manually to a nonverbal probe in Experiment 1.

Table 8 reveals that recall performance was significantly higher for the incidental-maintenance condition than for the intentional-elaborative condition, $F(1, 42) = 4.39$. Recall performance also dropped significantly with the increasing number of rehearsals, $F(2, 84) = 12.71$. There was no interaction effect of these two variables, $F(2, 84) = 2.11$. All of these trends were expected and have been discussed previously.

Word recognition. Table 9 presents the percentage of correct recognition on the final recognition test. These results nicely replicate those from Experiment 1. The Rehearsal Type × Rehearsal Interval interaction (for intervals 2 vs. 4) was not reliable, $F(1, 42) = 1.29$, $MS_e = 51.02$, whereas the comparable interaction for intervals 4 versus 10 was reliable, $F(91, 42) = 5.51$, $MS_e = 62.17$. These results, like those of Experiment 1, are consistent with the hypothesis that elaborative rehearsal has a monotonic effect on long-term recognition performance within the range explored, an effect that is not limited to the first few rehearsals as is the case with maintenance rehearsal.

It is interesting to note that unlike the number recall task, word recognition performance was as good in this experiment as it was in Experiment 1. It may be that long-term memory performance (word recognition) is not as affected by interference from the probe as short-term performance (number recall) may be.

Probe responses. We operationalized resistance to interruption in two ways. The first measure of susceptibility to interruption was the probability of actually inhibiting rehearsal at each rehearsal interval for each rehearsal condition. The results of this measurement appear in Table 10. Overall, there was no reliable difference between the two rehearsal conditions, $F(1, 42) = 3.11$, $MS_e = 582.1$. We calculated planned comparisons for the Rehearsal Type × Rehearsal Interval interaction (for the 2 vs. 4 and 4 vs. 10 data) in Table 10. The interaction was not significant for intervals 2 versus 4, $F(1, 42) = 1.31$, $MS_e = 58.3$. However, it was significant for intervals 4 versus 10, $F(1, 42) = 4.32$, $MS_e = 66.1$. This analysis, then, supports the prediction of our two-stage model: There was a marked decrease in probability of interruption in the maintenance rehearsal condition from 4 to 10 rehearsals, but there was no effect from 2 to 4 rehearsals. By this we conclude that interruption became more difficult during the second stage of maintenance rehearsal.

The second analysis of susceptibility to interruption examined the time subjects took to start responding to the probe word after the probe was presented. All sessions were recorded on audio tapes and were analyzed offline using a Mingograph machine. (As in Experiment 1, the scoring of the Mingograph output was done by two judges who were blind to the hypotheses of the experiment. The scoring of the judges agreed extremely well.) By examining the speech waveform of the probes and rehearsal utterances, we were readily able to measure the latency of the beginning of the probe utterance from its time of appearance. We accomplished this measurement in the following way: The time of appearance of the probe after the tone was known to be 175 ms (see *Method* section). So we measured the time from tone appearance (which could be clearly detected in the Mingograph output) to the point at which the probe word was uttered. From this value, we subtracted 175 ms to obtain the RT from probe appearance to response.

Although subjects generally complied with the instructions of the experiment, we separated their probe responses into two categories.

Table 10
Probability of Interrupting Rehearsal as a Function of Number of Rehearsals and Rehearsal Condition

Rehearsal condition	No. of rehearsals 2	*SD*	4	*SD*	10	*SD*
Incidental maintenance	.64	.32	.61	.30	.49	.28
Intentional elaborative	.56	.32	.53	.30	.54	.29

One category was characterized by subjects actually interrupting rehearsal to voice the probe (56% of the trials). The other type was defined as completion of the rehearsal before uttering the probe word (44% of the trials). These proportions were similar in the two conditions. Because these two different modes of response might involve different processes, they were analyzed separately. (Although it still might be the case that for the second type of response subjects were interrupting their processing, the first type is a better behavioral manifestation of interruption.) These two different modes of response to the probes could at least partially be due to a differential interpretation of the instructions. This is suggested by the fact that half of the subjects ($n = 22$) had one of the two types of response predominating (i.e., they had 90% or more of their responses being either rehearsal completions or rehearsal interruptions). Such an extreme distribution suggests that we are dealing with two strategies. (Because not all subjects produced the same proportion of trials of each response type, subjects were included in the analysis only if they had at least 40% of their total responses of a given response type.)

Table 11 presents RTs to the probe as a function of rehearsal type and number of rehearsals on those trials in which subjects actually interrupted rehearsal.

In order to evaluate predictions of the two-stage model of maintenance rehearsal, the planned comparisons described for the word recognition data of this experiment were used for the present data as well.

The Rehearsal Type × Rehearsal Interval interaction was not reliable for the 2 versus 4 comparison, $F(1, 34) = 2.01$, $MS_e = 787.3$. However, the comparable interaction for the

Table 11
Mean Vocal Probe Reaction Times (in ms) as a Function of Number of Rehearsals and Rehearsal Condition

Rehearsal condition	No. of rehearsals[a] 2	*SD*	4	*SD*	10	*SD*
Incidental maintenance	620	55	617	43	646	48
Intentional elaborative	610	59	600	54	608	52

[a] Data presented are for trials on which subjects overtly interrupted the rehearsal sequence.

4 versus 10 interval was statistically significant, $F(1, 34) = 4.51$; $MS_e = 730.2$.

The obtained results of probe RT for both rehearsal conditions are as predicted. These results support the view that maintenance rehearsal becomes more automated with time, and therefore more difficult to interrupt. Beyond the significance of the results, it is important to note their direction in comparison to Experiment 1. Although there was no change in the direction of the results in the two experiments in the elaborative rehearsal condition, the direction for the maintenance rehearsal condition (between 4 and 10 rehearsals) is just the opposite to the one obtained in Experiment 1. It supports the use of the two different probe tasks chosen to operationalize the different criteria for automaticity.

Now let's turn to the other category of probe responses. Table 12 presents probe RT as a function of rehearsal condition and number of rehearsals on those trials in which subjects completed the rehearsal sequence before they responded to the probe. Several points should be noted about the results. First, the overall RT was slower by about 100 ms for these responses compared with the first type. Such a trend is to be expected, because it took subjects extra time to complete the rehearsal sequence before responding. Second, neither the interaction between 2 and 4 rehearsals and rehearsal type nor that between 4 and 10 rehearsals and rehearsal type proved to be significant.

Although not significant, note the decrease in RT from 4 to 10 rehearsals in the maintenance rehearsal condition. For those trials in which the subjects used this mode of response (first completing the rehearsal and then responding to the probe), they had to process the probe while they were rehearsing the pair of words. From their point of view it might have been a dual task. If that was the case, then the decrease in probe RT from 4 to 10 rehearsals may be similar to the comparable effect of Experiment 1.

There is a final point about the difference in probe RT between the two rehearsal conditions for this type of response. The probe RT in the maintenance rehearsal condition was higher by about 80 ms than in the elaborative rehearsal condition. If anything, we would have expected the opposite result, because subjects in the elaborative rehearsal condition presumably devoted more attention to rehearsal, hence leaving them with less capacity devoted to the probe task. The reason for this difference, however, lies in a subject-selection problem. Only 13 of the 21 subjects in the maintenance rehearsal condition participated in the analysis of this type of trial. These subjects proved to be slower than the others in their overall RTs (by about 80 ms), accounting for the differences between the two rehearsal conditions.

Trade-off analysis. In order to assess trade-off possibilities between the number and the rehearsal tasks, log odds ratios were calculated as in Experiment 1. The values from this calculation transformed back to odds ratios are shown in Table 13.

For the incidental-maintenance condition, no odds ratios were significantly different from 1.0 or from each other, indicating no changes in trade-offs across the rehearsal intervals. For

Table 12
Mean Vocal Probe Reaction Times (in ms) as a Function of Number of Rehearsals and Rehearsal Condition

Rehearsal condition	No. of rehearsals[a] 2	*SD*	4	*SD*	10	*SD*
Incidental maintenance	790	59	787	64	772	61
Intentional elaborative	722	64	713	67	704	65

[a] Data presented are for trials on which subjects did not overtly interrupt the rehearsal sequence.

Table 13
Mean Odds Ratios of Recall Versus Recognition Measures as a Function of Number of Rehearsals and Rehearsal Condition

Rehearsal condition	No. of rehearsals 2	4	10
Incidental maintenance	0.99	1.09	1.06
Intentional elaborative	1.01	0.87	1.14

the intentional-elaborative condition, the odds ratio for the rehearsal interval of 4 was significantly smaller than 1.0, and it was smaller than the ratios for the other two rehearsal intervals. This raises the possibility that at least part of the increase in word recognition performance from 2 to 4 rehearsals in the intentional-elaborative condition was due to a decrease in number recall performance. But because the increase in recognition performance from 4 to 10 rehearsals in this condition was not associated with a decrease in number performance (odds ratios increased from 4 to 10 rehearsals and were larger than 1.0), it rules out the possibility that the entire recognition effect in this condition was due to shifts in allocation policies.

Taken together, the results of this experiment support the proposed model of maintenance rehearsal. The word recognition results replicated the pattern obtained in Experiment 1, in which only the first few rehearsals enhanced long-term recognition performance. The probe RT data support the notion that the second stage in maintenance rehearsal is more automatic as revealed by its being more difficult to suppress.

General Discussion

Let us review the major features of the model of maintenance rehearsal that we have proposed. We claim that maintenance rehearsal is composed of two stages: The first involves retrieval of an appropriate motor program for a rehearsal sequence and initiation of the execution of this program. The second stage involves continuous execution of this motor program. By hypothesis, the first stage is more demanding of cognitive resources than is the second, and it is more firmly under the subjects' control. This claimed difference in automaticity between the stages is largely supported by the results of the experiments just reported: Compared with the first stage, the second demands less in the way of processing capacity; it results in more stereotyped rehearsal utterances, and it is more difficult to interrupt.

A major feature of the model is that it allows one to make a prediction about the later memory consequences of maintenance rehearsal. The prediction is that Stage 1 of rehearsal would produce traces in memory that would endure well beyond the rehearsal itself. Stage 2, being more automatic, would have relatively little effect on the creation of long-term traces.

The rationale for this prediction rests on previous research that reveals an effect of cognitive effort on the storage of information in memory. The greater the effort or cognitive capacity devoted to a memory task, regardless of the level of processing employed, the greater is the probability of later retrieving the material that was studied (Eysenck & Eysenck, 1979; Tyler et al., 1979). This cognitive effort hypothesis has been linked to the effects of repetition on retrieval by Johnston and Uhl (1976). These authors directly tested an effort theory of the spacing effect (Hintzman, 1974). They predicted that memory for an item is positively correlated with the total amount of processing accorded the item. They reasoned that after a first presentation, an item receives a base level of processing effort. The processing effort accorded the second presentation of the item was assumed to vary with the length of the spacing interval, being at some point below base level at a short interval and rising toward base level as the interval increases. By measuring the processing effort extended to each encoding of an item through use of a dual-task paradigm, they found that encoding effort was an accurate predictor of recall accuracy, in general, and repetition spacing, in particular. Probe RT on the secondary task increased somewhat in the spaced repetition condition and decreased in the massed repetition condition. These results are similar to the ones obtained in our first experiment in which probe RT increased somewhat in the elaborative rehearsal condition and decreased in the maintenance rehearsal condition. The effects on long-term memory were also similar, at least qualitatively.

The empirical similarities between the spacing effect and maintenance rehearsal should be expected considering the fact that both phenomena are cases of repetition effects on memory (Craik & Jacoby, 1979). Being able to explain both effects with the concept of automaticity of processing (or of effort expended) may encourage further research on the parallel between the phenomena (for a related point see Nairne, 1983; and for some potentially conflicting data, see Krinsky & Nelson, 1981).

We may well ask at this point how added effort may enhance memory for an item. A potential answer to this question rests on a distinction proposed by Mandler (1979). He distinguished between two dimensions of memory organization: integration and elaboration. Integration is presumed to enhance an item's intraunit cohesiveness, whereas elaboration refers to the building of interitem connections. We propose that prior to and during the first few rehearsals in a sequence, information is retrieved from permanent memory, an articulatory loop is constructed, and it is executed a few times. This may increase the integration of the pair of words in the loop (intraunit cohesiveness mostly but perhaps with some interitem connection as well) and hence may lead to an increase in recognition performance. We should note that Mandler proposed that an increase in integration is associated with a reduction of the cognitive capacity devoted to an item. This reduction, which presumably occurs in Stage 2 when the rehearsal program is being executed repetitively, is then associated with no further increase in long-term recognition performance.

Let us return for a moment to the controversy over whether maintenance rehearsal has any effect on later retrieval. Our data, and those of others, reveal that the proper circumstances can produce such an effect (Geiselman & Bjork, 1980; Maki & Schuler, 1980; Nairne, 1983; Naveh–Benjamin and Jonides, in press; Woodward et al., 1973). This contrasts with many reports failing to find such an effect, as we have cited earlier. As we discussed, part of the discrepancy may be attributable to the use of recall versus recognition tests. Perhaps recall is simply not sufficiently sensitive to measure the subtle effects of maintenance rehearsal, especially if these effects are confined to the first several rehearsals, as our data and those of Geiselman and Bjork (1980), Nairne (1983), and Naveh–Benjamin and Jonides (in press) indicate. Another factor in these experiments may be the extent to which the primary task taxes processing capacity, thereby leaving minimal capacity for maintenance rehearsal. Many experiments using the paradigm that we adopted have provided subjects with a primary memory task that might have been insufficiently taxing (e.g., Glenberg & Adams, 1978; Rundus, 1980). The effect of this would be to free up processing resources for rehearsal, perhaps distorting the effect that this rehearsal may have on later retrieval. A final contribution to the discrepancy in the literature has to do with the particular effect that maintenance rehearsal has on memory. Our data, but even more convincingly the data of Nairne (1983) and Glenberg and Bradley (1979), have led us to the view that maintenance rehearsal enhances intraunit cohesiveness, as we previously discussed. During the first part of a rehearsal cycle, when articulatory codes are being retrieved and when execution begins, the cognitive effort involved in this activity increases the intraunit bond. In our case, this amounts to a strengthening of not only the familiarity of the words themselves but also of the link between them. This in turn facilitates retrieval of word pairs later on.

The model that we have proposed here bears an interesting resemblance to the concept of working memory as discussed by Baddeley (1976) and Baddeley and Hitch (1974). According to this view, short-term memory is best viewed as a combination of a central processor and an articulatory loop. The processor can be responsible for such activities as retrieval from permanent memory and recoding, whereas the loop is simply a recirculatory mechanism. It seems plausible to identify the first stage of our model as the central processor and the second stage as the articulatory loop of working memory. Whether this parallel is illuminating, however, would have to be a matter for future research. It is interesting to note a recent related discussion by Nairne (1983) in which he raises the possibility that the beginning of the rehearsal interval creates associations between the to-be-rehearsed items. This would presumably be a function of central processor activity.

Before concluding, let us make two methodological remarks about our experiments. The first concerns the strength of our argument about a relationship between stages of maintenance rehearsal and their memory consequences. The potential problem here is that we did not directly manipulate the amount of attention devoted to each stage—we measured it after the fact. Given this correlational characteristic of our design, one may feel that we are not entitled to draw a causal link between the effort expended in each stage and the later recognition results. It was for precisely this reason that we developed the reported converging sources of evidence that would strengthen the case for the involvement of automaticity.

The second point concerns the relatively restricted range of rehearsal intervals used (1–10). Although in principle the range could be extended beyond 10 rehearsals, in practice there are problems with this experimental strategy. It seems likely that with an extended period of rehearsal, subjects may begin to suspect a later memory test (perhaps this is the cause of Rundus's, 1980, effects on free recall after 60 s of maintenance rehearsal). Also, such extended periods of repetition do not typify the everyday use of maintenance rehearsal. It is for these reasons that we chose a restricted range.

By way of concluding, let us ask what the implications of our results are for the two currently popular theories of memory: the structures theory and the levels-of-processing view. We and others have established that maintenance rehearsal does have an effect on long-term memory, in direct contradiction to a prediction of the levels-of-processing notion. But our results go beyond this in showing that even the seemingly simple control process of maintenance rehearsal is more complex than previously expected. Perhaps it is the more automatic second stage of maintenance rehearsal that the architects of the levels-of-processing framework had in mind when making their prediction. Perhaps it is the first stage of maintenance rehearsal that authors of the structures view have focused on in claiming that processes such as maintenance rehearsal are under subject control and have long-term memory effects. It should be disconcerting to both camps to discover that mere repetition is not an elementary process. Perhaps other control processes deserve the same sort of scrutiny.

References

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 2, pp. 89–105). New York: Academic Press.

Baddeley, A. D. (1976). *The psychology of memory.* New York: Basic Books.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. A. Bower (Ed.), *The psychology of learning and motivation* (pp. 47–90). New York: Academic Press.

Craik, F. I. M., Gardiner, J. M., & Watkins, M. J. (1970). Further evidence for a negative recency effect in free recall. *Journal of Verbal Learning and Verbal Behavior, 9,* 554–560.

Craik, F. I. M., & Jacoby, L. L. (1979). Elaboration and distinctiveness in episodic memory. In L. G. Nilsson (Ed.), *Perspectives on memory research: Essays in honor of Uppsala University's 500th anniversary* (pp. 145–166). Hillsdale, NJ: Erlbaum.

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

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

Dark, V. J., & Loftus, G. R. (1976). The role of rehearsal in long-term memory performance. *Journal of Verbal Learning and Verbal Behavior, 15,* 479–490.

Darley, C. F., & Glass, A. L. (1975). Effects of rehearsal and serial list position on recall. *Journal of Experimental Psychology: Human Learning and Memory, 104,* 453–458.

Eysenck, M. W., & Eysenck, M. C. (1979). Processing depth, elaboration of encoding, memory stores, and expanded processing capacity. *Journal of Experimental Psychology: Human Learning and Memory, 5,* 472–484.

Geiselman, R. E., & Bjork, R. A. (1980). Primary versus secondary rehearsal in imagined voices: Differential effects on recognition. *Cognitive Psychology, 12,* 188–205.

Geiselman, R. W., Woodward, J. A., & Beatty, J. (1982). Individual differences in verbal memory performance: A test of alternative information-processing models. *Journal of Experimental Psychology: General, 111,* 109–134.

Glenberg, A., & Adams, F. (1978). Type I rehearsal and recognition. *Journal of Verbal Learning and Verbal Behavior, 17,* 455–463.

Glenberg, A., & Bradley, M. M. (1979). Mental contiguity.

Journal of Experimental Psychology: Human Learning and Memory, 5, 88–97.

Glenberg, A., Smith, S. M., & Green, C. (1977). Type I rehearsal: Maintenance and more. *Journal of Verbal Learning and Verbal Behavior, 16,* 339–352.

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

Hintzman, D. L. (1974). Theoretical implications of the spacing effect. In R. L. Solso (Ed.), *Theories in cognitive psychology: The Loyola Symposium* (pp. 77–100). Potomac, MD: Erlbaum.

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

Johnston, W. A., & Heinz, S. P. (1978). Flexibility and capacity demands of attention. *Journal of Experimental Psychology: General, 107,* 420–435.

Johnston, W. A., & Uhl, C. N. (1976). The contributions of encoding effort and variability to the spacing effect on free recall. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 153–160.

Jonides, J. (1981). Voluntary versus automatic control over the mind's eye's movement. In J. Long & A. D. Baddeley (Eds.), *Attention and performance IX* (pp. 187–204). Hillsdale, NJ: Erlbaum.

Keppel, G. (1982). *Design and analysis: A researcher handbook* (2nd ed.), Englewood Cliffs, NJ: Prentice-Hall.

Krinsky, R., & Nelson, T. O. (1981). Task difficulty and pupillary dilation during incidental learning. *Journal of Experimental Psychology: Human Learning and Memory, 7,* 293–298.

Kučera, H., & Francis, W. (1967). *Computational analysis of present-day American English.* Providence, RI: Brown University Press.

Logan, G. D. (1979). On the use of a concurrent memory load to measure attention and automaticity. *Journal of Experimental Psychology: Human Perception and Performance, 5,* 189–207.

Logan, G. D. (1981). Attention, automaticity, and the ability to stop a speeded choice response. In J. Long & A. D. Baddeley (Eds.), *Attention and performance IX,* (pp. 205–222). Hillsdale, NJ: Erlbaum.

MacKay, D. G. (1982). The problems of flexibility, fluency and speed-accuracy trade-off in skilled behavior. *Psychological Review, 89,* 483–506.

Maki, H., & Schuler, J. (1980). Effects of rehearsal duration and levels of processing on memory for words. *Journal of Verbal Learning and Verbal Behavior, 19,* 36–45.

Mandler, G. D. (1979). Organization and repetition: An extension of organizational principles with special reference to rote learning. In L. G. Nilsson (Ed.), *Perspectives on memory research: Essays in honor of Uppsala University's 500th anniversary* (pp. 293–327). Hillsdale, NJ: Erlbaum.

Mechanic, A. (1962). The distribution of recalled items in simultaneous intentional and incidental learning. *Journal of Experimenal Psychology, 63,* 593–600.

Meunier, G. F., Kestner, J., Meunier, J. A., & Ritz, D. (1974). Overt rehearsal and long-term retention. *Journal of Experimental Psychology, 102,* 913–914.

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

Modigliani, V. (1976). Effects of delaying an initial recall on a later recall. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 609–622.

Modigliani, V., & Seamon, J. G. (1974). Transfer of information from short-to-long-term memory. *Journal of Experimental Psychology, 102,* 768–772.

Nairne, J. S. (1983). Associative processing during rote rehearsal. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9,* 3–20.

Naveh–Benjamin, M., & Jonides, J. (in press). Cognitive load and maintenance rehearsal. *Journal of Verbal Learning and Verbal Behavior.*

Nelson, T. O. (1977). Repetition and depth of processing. *Journal of Verbal Learning and Verbal Behavior, 16,* 151–171.

Palmer, J. C., Jonides, J. & Palmer, C. (1980, November). *Lack of voluntary control as a criterion for automatic processing.* Paper presented at the meeting of the Psychonomic Society, St. Louis, MO.

Posner, M. I., & Snyder, C. R. (1975). Attention and cognitive control. In R. Solso (Ed.), *Information processing and cognition: The Loyola Symposium* (pp. 55–86). Hillsdale, NJ: Erlbaum.

Roenker, D. L. (1974). Role of rehearsal in long-term memory. *Journal of Experimental Psychology, 103,* 368–371.

Rundus, D. (1971). Analysis of rehearsal processes in free recall. *Journal of Experimental Psychology, 89,* 63–77.

Rundus, D. (1977). Maintenance rehearsal and single-level processing. *Journal of Verbal Learning and Verbal Behavior, 16,* 665–681.

Rundus, D. (1980). Maintenance rehearsal and long-term recency. *Memory & Cognition, 8,* 220–230.

Schneider, W., & Shiffrin, R. N. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84,* 1–66.

Stanners, R. F., Meunier, G. F., & Headley, D. B. (1969). Reaction time as an index of rehearsal in short-term memory. *Journal of Experimental Psychology, 82,* 566–570.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18,* 643–662.

Tulving, E. (1966). Subjective organization and the effects of repetition in multi-trial free recall verbal learning. *Journal of Verbal Learning and Verbal Behavior, 5,* 193–197.

Tyler, S. W., Hertel, P. T., McCallum, M. C., & Ellis, H. C. (1979). Cognitive effort and memory. *Journal of Experimental Psychology: Human Learning and Memory, 5,* 607–617.

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

Received June 24, 1983
Revision received February 16, 1984 ■