

Cognitive Load and Maintenance Rehearsal

MOSHE NAVEH-BENJAMIN

Ben-Gurion University of the Negev, Israel

AND

JOHN JONIDES

University of Michigan

In recent years there has been a good deal of debate about the role of rote, repetitive rehearsal (called Type I or maintenance rehearsal) on the establishment of memory traces that outlast the rehearsal process itself. One advance in the technology used to study this problem is the operational definition of maintenance rehearsal proposed by Glenberg and Adams (1978, *Journal of Verbal Learning and Verbal Behavior*, 17, 455-463). These authors argued that maintenance rehearsal should be defined as the continuous maintenance of information in memory using minimal cognitive capacity. Here this definition was adopted and extended in a paradigm in which the mental resources devoted to maintenance rehearsal could be systematically varied. The experiment revealed that there is, indeed, an effect of maintenance rehearsal on long-term recognition performance and that this effect depends on the mental resources devoted to the rehearsal process.

The functions of rehearsal are two in number: to maintain information in a temporarily active state during short-term tasks and to create memory traces with some permanence. The latter of these functions has often been attributed to the class of processes called "elaborative" rehearsal, processes such as chunking, forming images, or recoding material in various ways. However, various sources of evidence have been accumulating to suggest that maintenance rehearsal—mainly rote repetition—may also play some role in creating longer term memories. If this finding is correct, it is significant for two reasons. From a practical point of view, rote repetition is the strategy that is probably used most frequently as a mnemonic, so it is important

to assess its efficacy in establishing memory traces with some permanence. From a theoretical point of view, the two most popular overall theories of memory differ in their claims about the role of repetition. The theory originally articulated by Atkinson and Shiffrin (1968) posits that maintenance rehearsal causes information to be transferred from short- to long-term memory. By contrast, according to the levels-of-processing view (Craik & Lockhart, 1972), maintenance rehearsal serves only to hold a particular level of code intact without creating a deeper level of code which, by the hypothesis, would result in a longer lasting trace.

The significance of this issue has resulted in a body of research on the long-term effects of maintenance rehearsal. In general, the results have been mixed. Some have found no connection between free recall and the amount of maintenance rehearsal (Craik & Watkins, 1973; Gotz & Jacoby, 1974; Jacoby & Bartz, 1972; Meunier, Ritz,

This research was supported in part by Grant BNS-8024512 from NSF, and in part by Grant 82-0297 from AFOSR. Requests for reprints should be addressed to J. Jonides, Department of Psychology, University of Michigan, 330 Packard Rd., Ann Arbor, MI 48104.

& Meunier, 1972; Modigliani & Seamon, 1974; Roenker, 1974; Rundus, 1977), while others have shown a direct relationship (Dark & Loftus, 1976; Darley & Glass, 1975; Maki & Schuller, 1980; Mechanic, 1962; Modigliani, 1976). Still others have found effects of maintenance rehearsal only on later recognition performance, and not recall (Woodward, Bjork, & Jongeward, 1973; Glenberg, Smith & Green, 1977).

As Nelson (1977) has pointed out, one reason for the ambiguity in results from these studies has to do with the definition of the maintenance process. Several investigators have defined maintenance rehearsal by its lack of effect on later recall, a definition that makes it impossible to study later memory effects of the process (e.g., Glenberg, Smith, & Green, 1977).

The solution to this problem is to adopt a definition of maintenance rehearsal that is independent of its effects on later memory tests. Such a definition was suggested by Glenberg and Adams (1978). They proposed to define it as a mnemonic process that "continuously maintains information in memory using the minimum amount of cognitive capacity necessary to maintain the information" (p. 456). To meet the condition of their definition, they developed a technique in which subjects were tested on memory for the distractors in a Brown-Peterson task using an incidental learning paradigm. Each trial consisted of the presentation of some numbers for study, followed by overt rehearsal of a word or words that served as the distractors for the number task. As in the typical Brown-Peterson task, subjects were required to recall the numbers on each trial. This was the primary task. The number of rehearsals required for the distractors was varied systematically to provide different amounts of rehearsal for this material. At the end of a series of these short-term memory trials, an unexpected recognition test for the rehearsed distractor words was administered. The rationale for using this task was that recall for the numbers served as the primary task to which subjects di-

rected attention. This made rehearsal of the words a task of secondary importance so that subjects would presumably devote minimal processing resources to its execution. Consequently, the final recognition test could assess the permanence of memory traces that may have been created by the minimally demanding rehearsal activity of repeating the words (a prototypical maintenance rehearsal strategy).

With the innovation of this new operational definition, Glenberg and Adams (1978) contrasted three different hypotheses about the effect of maintenance rehearsal on later recognition of the rehearsed words:

(1) the "tagging" hypothesis, according to which a relationship between rehearsal and recognition is attributed to the addition of frequency or context tags to the memory trace,

(2) the "depth" hypothesis, which attributes the improvement in recognition with additional rehearsals to the accumulation of semantic information about the memory trace,

(3) the "acoustic" hypothesis, according to which rehearsal and recognition are related because rehearsal causes acoustic information about a trace to accumulate.

Using a specially designed recognition test (described later under Method), they found that maintenance rehearsal does play a role in establishing long-term memory traces. Specifically, it appears to strengthen the acoustic-phonemic components of the rehearsed information. They concluded that their findings were in conflict with the original Type I-Type II rehearsal distinction suggested by Craik and Lockhart (1972).

Since these conclusions depend on the operational definition of Type I (maintenance) rehearsal, it is essential to examine its adequacy. Let us consider two key issues that are germane to this examination.

Minimizing capacity. The crucial feature of the present operational definition of maintenance rehearsal is that the process must make use of only the *minimum*

amount of cognitive capacity necessary to maintain information temporarily. Glenberg and Adams tried to establish this by having rehearsal be an incidental task. But this may not have been sufficient. It is possible that the number recall task was not as difficult as necessary to command control over most of the processing capacity available. This may have happened, for example, if subjects developed chunking strategies that allowed them to reduce the four single-digit numbers to fewer than four chunks of information, making the numbers less difficult to retain even with interference. Judging from the data in their experiment, this seems a reasonable hypothesis. The *average* accuracy for numbers in their experiment was around 84% for 2, 6, and 10 rehearsals, not very far from ceiling. Moreover, after only one rehearsal, performance was no higher than 85%, so further rehearsals must barely have changed recall performance at all. This pattern of results suggests that the numbers might have been secured in memory after 1 rehearsal, then subjects may have switched their attention to the word task, hence using more than the minimal capacity necessary for mere maintenance.

Measuring capacity. There is a general problem raised by the discussion of minimizing capacity: How can one measure whether subjects trade resources on each trial between the act of rehearsing and the primary task of retaining the numbers?

Glenberg and Adams (1978) suggested that evidence of no change in number recall with additional rehearsals would be evidence for a constant allocation of capacity to rehearsal. But this is an unreasonably stiff criterion. Although the issue of the forgetting agent (decay or interference) in short-term memory (STM) is not yet settled, the assumption underlying this criterion is that *no* decay occurs over time, or that interference by the rehearsed words does *not* cause any decrement in primary task performance. But in a Brown–Peterson task, with backward counting as the

interpolated activity, typical results show a decrement in recall over time, even though no increase in cognitive capacity devoted to the secondary task is assumed (see Rundus, 1980, for a related argument). So we should be skeptical of results showing *no change* in recall performance with additional rehearsals because *most* reasonable theories of forgetting would predict a decrease. In fact, one might interpret no change in recall as indicating a decrease in cognitive capacity devoted to additional rehearsals. Thus, one might argue that the reason for Glenberg and Adams' not finding support for the "depth hypothesis" was a decrease in cognitive capacity allocated to additional rehearsals.

The conclusion to draw from this line of argument is that merely using recall performance as a diagnostic of resource allocation is not particularly reasonable. There are two alternative methods that are superior.

On a macroscopic level, one could compute a correlation between primary task performance and later recognition of the rehearsed words, across different lengths of the rehearsal period. A trade-off between resources devoted to the primary task and to rehearsal would be indicated by a negative correlation. That is, as the number of rehearsals increased, if subjects traded resources between number memorization and rehearsal, a decline in number recall would be accompanied by an increase in word recognition. If obtained, this correlation would complicate an interpretation of the relationship between number of rehearsals and later recognition. The chief problem with this sort of gross analysis, however, is that it is susceptible to the error known as Simpson's paradox. Even with no overall negative correlation present across trials, there might still be a trial-by-trial relationship between recall and recognition scores (see Hintzman, 1980, for a complete discussion of this problem in the context of other issues in memory research).

There is a more powerful variant of the

correlational logic that provides a microscopic diagnosis of the same trade-off problem by analyzing data on a trial-by-trial basis, minimizing the effect of Simpson's paradox. Suppose that subjects do trade resources between number memorization and rehearsal of the distractor words. Then a trial on which a subject concentrated heavily on retaining the numbers in memory should result in poor later recognition of the rehearsed words, and vice versa. This prediction can be readily tested by creating a contingency table that relates recall of the numbers on any given trial to probability of later recognition. A trade-off in a contingency table of this sort would be revealed by calculating an odds ratio between recall and recognition. To see why this is so, examine the mock data displayed in Table 2. The two cells along the major diagonal in both the left and right halves of this table represent trials on which subjects were correct on either recall or recognition tests, but not on both. The values in these cells ought to be large if subjects were trading resources between number retention (which presumably is related to number recall) and rehearsal (which, by the hypothesis, is related to distractor recognition). An odds ratio would reveal such a trade-off since it is simply the product of the minor diagonal divided by the product of the major diagonal. To the extent that this ratio is less than 1.0, it reveals a trade-off between tasks. If the ratio is 1.0, the cells are independent, and there is no trade-off indicated.

The purpose of this sort of analysis is not only to discover whether subjects trade recall accuracy for later recognition accuracy, but also to discover whether this trade-off can account for an increase in later recognition with increased rehearsals. To meet this second goal, one must partition the data of an experiment such as Glenberg and Adams' 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 the number of rehearsals. It is this last analysis that will reveal whether a trade-off has contaminated the effect of number of rehearsals on later recognition.

Let us examine two hypothetical examples to concretize this argument. Consider first the example illustrated in Table 1. These hypothetical data could have been drawn from a Brown-Peterson task in which subjects were required to recall numbers on each trial after they had rehearsed distractor words for one of two rehearsal intervals, X or $X + Y$. In this experiment, furthermore, after all trials were completed, subjects were given a surprise recognition test of the rehearsed words. The table displays hypothetical recall and recognition performance as a function of rehearsal interval. On the face of it, one might argue that the increase in recognition performance from 40 to 50% with increased number of rehearsals is simply a function of a trade-off with the decline in recall from 80 to 60%.

But the contingency data in Table 2 suggest that this is not the case. The table presents contingency data on the relationship between recall and recognition for each number of rehearsals. As you can see, calculation of the odds ratios for each number of rehearsals suggests that recall and rehearsal do not trade-off with one another (odds ratio = 1.0 for each). From this, it is extremely cumbersome and implausible to argue that while there is no trade-off for

TABLE 1
A HYPOTHETICAL EXAMPLE OF MEAN PERCENTAGES OF CORRECT RECALL AND RECOGNITION RESPONSES FOR TWO DIFFERENT NUMBER OF REHEARSALS

Measure	Rehearsal interval	
	X	$X + Y$
Recall performance	80	60
Recognition performance	40	50

TABLE 2
A HYPOTHETICAL EXAMPLE OF AN ASSOCIATION BETWEEN RECALL AND RECOGNITION PERFORMANCE FOR DIFFERENT NUMBERS OF REHEARSALS

	Rehearsal interval					
	X			X + Y		
	Recognized	Not recognized	Total	Recognized	Not recognized	Total
Recalled	32	48	80	30	30	60
Not recalled	8	12	20	20	20	40
Total	40	60	100	50	50	100
Odds ratio	1.00			1.00		

each interval alone, there is nevertheless a trade-off across intervals. Rather, we take such a result to indicate that the increase in recognition performance across intervals is not simply a function of the drop in recall.

Now consider a second example illustrated in Table 3. Here we see an increase in recognition accuracy across rehearsal interval (40% to 50%) with no change in recall (80% vs 80%). One might be tempted in this case to argue that the change in recognition performance can be attributed only to a change in number of rehearsals, and not to any difference in recall performance.

Further analysis suggests that this argument would be premature, however. Table 4 displays a hypothetical contingency tabulation of the data in Table 3. For the first number of rehearsals, X , we see that subjects appear to perform as if the recall and recognition tasks are independent of one another. But for trials with $X + Y$ re-

hearsals, there is a definite trade-off between recall and recognition: Subjects show an increase in recognition at the expense of their recall scores. Thus, one might argue from these data that the increase in overall recognition from 40 to 50% is to some extent attributable to a change in subjects' trade-off posture with increased rehearsals. Consequently, one should be cautious in this case about drawing conclusions about a direct relationship between rehearsals and recognition.

These examples can be summarized as follows: The first case shows that a change in recall performance is not sufficient to infer a change in allocation policy of cognitive resources across number of rehearsals. The second case shows that it is also not necessary. Instead, we advocate a more microscopic analysis of the trial-by-trial data themselves, an analysis that can be accomplished through the use of contingency tables that include recall and recognition scores.¹ This analysis minimizes problems due to Simpson's paradox by separately evaluating a potential trade-off for each subject and each condition.

Turning to our empirical work, the major purpose of the present experiment was to

TABLE 3
A SECOND HYPOTHETICAL EXAMPLE OF MEAN PERCENTAGE OF CORRECT RECALL AND RECOGNITION RESPONSES FOR TWO DIFFERENT NUMBER OF REHEARSALS

Measure	Rehearsal interval	
	X	X + Y
Recall performance	80	80
Recognition performance	40	50

¹ Note one limitation of this analysis. Whenever the values on one of the variables (recall or recognition) are close to the floor or ceiling of performance, there will be a small number of observations in one or more cells of the corresponding contingency table. This will render the representativeness of the odds ratio suspect.

TABLE 4
A SECOND HYPOTHETICAL EXAMPLE OF AN ASSOCIATION BETWEEN RECALL AND RECOGNITION PERFORMANCE FOR DIFFERENT NUMBERS OF REHEARSALS

	Rehearsal interval					
	X			X + Y		
	Recognized	Not recognized	Total	Recognized	Not recognized	Total
Recalled	32	48	80	32	48	80
Not recalled	8	12	20	18	2	20
Total	40	60	100	50	50	100
Odds ratio	1.00			0.07		

refine the technology developed by Glenberg and Adams (1978) to study maintenance rehearsal. The procedural innovation that we introduced was a variation in the cognitive load demanded by the primary task. This was accomplished by varying the number of 2-digit numbers that subjects were required to retain (two, three, or four) on each of the Brown-Peterson trials. The effect of this variation was presumably to cause a concomitant variation in the capacity available for rehearsing the distractor words. This procedural refinement permitted us to conduct three tests concerning the effect of maintenance rehearsal:

(a) We were able to attempt a replication of the results of Glenberg and Adams since the lightest level of memory load of the primary task (two 2-digit numbers on each trial) was quite similar to their conditions.

(b) We were able to provide a contingency analysis of recall and recognition across primary task loads and number of rehearsals. Even more importantly, the task allowed a rigorous test of the possibility that changes in recognition with increasing rehearsals are due to a changing trade-off of resource allocation between retention of the numbers and rehearsal.

(c) By varying the demand of the primary task, it became possible to test whether minimal capacity was available for rehearsal. This is crucial since the noncircular definition of maintenance rehearsal

introduced by Glenberg and Adams (1978) uses minimal capacity as the hallmark.

What results should one expect for the effect of number of rehearsals on later incidental recognition if primary task load is varied? Several predictions are available from current theories: First of all, a direct reading of the levels-of-processing approach predicts that maintenance rehearsal will have no effect in laying down memory traces of the rehearsed words for later recognition, regardless of primary task load. By contrast, the approach to memory originally articulated by Atkinson and Shiffrin (1968) predicts a positive effect of rehearsals on later recognition. But in this case, as more capacity is drawn away from the rehearsal process by increasing primary task load, the effect of rehearsals on recognition should diminish.

Finally, we might ask about the *mechanism* underlying an effect of rehearsals on recognition. Two possibilities suggest themselves. One is that increased maintenance rehearsal may simply increase the strength of a trace without affecting the nature of the code in which the trace is stored. According to this hypothesis, an increase in capacity devoted to rehearsal will simply increase later recognition performance. A second possibility is that maintenance rehearsal causes memory traces to change in code to a different and more permanent form much as elaborative rehearsal is as-

sumed to cause (e.g., acoustic to semantic). This hypothesis then predicts that increased capacity devoted to rehearsal will result not only in increased recognition, but also in evidence that the nature of the stored memory trace has changed. Our experiment was designed to discriminate between these alternatives.

METHOD

Subjects. The subjects were 45 undergraduate students, male and female, at Ben-Gurion University of the Negev. They were paid \$2.00 for their participation in a 75-minute session.

Design. The basic task was modeled after the Brown-Peterson procedure. Subjects were first presented with two, three, or four 2-digit numbers to hold in memory. Then they rehearsed a pair of words for a period of time that varied across trials (The particular words used on each trial differed, and the words were counterbalanced across subjects for the number of times each was rehearsed). Finally, subjects were required to write down the numbers that they had memorized on each trial. At the end of the experiment, they were given an unexpected test of their recognition for the words.

The difficulty of the primary task load (two, three, or four 2-digit numbers) was manipulated across subjects, with 15 subjects assigned to each task load.

Altogether, there were 86 trials in the experiment. Of these, the first 6 were practice. The next 4 trials and the last 4 were included to absorb primacy and recency effects on the later recognition test (for a similar procedure see Glenberg & Adams, 1978; Glenberg, Green, & Smith, 1977; and Rundus, 1980). Data from the remaining 72 trials were included in the analyses presented below.

The 72 test trials were composed of 24 for which subjects engaged in 1 rehearsal, 24 for which 5 rehearsals were required, and 24 for which 10 rehearsals were required. The number of rehearsals required on any trial was not predictable.

Apparatus and materials. All stimuli were presented visually via an externally timed Kodak Carousel slide projector, controlled by one channel of a stereo tape deck.

The numbers used for the primary task on each of the 86 trials consisted of two, three, or four 2-digit numbers. These sets of numbers were selected randomly with the following constraints: (a) All numbers were drawn from the numbers 12 to 98 inclusive. (b) No numbers whose two digits were identical (e.g., 55) were included. (c) No two numbers in a set were integer multiples of one another.

The pair of words used for rehearsal on each trial consisted of one- or two-syllable common Hebrew words. There were no obvious semantic or acoustic relationships between the words in each pair.

The final recognition test was a five-alternative forced-choice task similar to the one used by Glenberg and Adams (1978). The recognition task was designed to assess the form of the code in which rehearsed items might be stored. For each of the 72 experimental trials, one word of each pair was chosen as the target word (T) on the recognition test (on half of the trials the top word, and on half, the bottom word on a slide). For each of the recognition trials, a set of four one- or two-syllable common Hebrew words was generated as a distractor set. One distractor for each trial was a synonym (S) or was closely related in meaning to the target. Another distractor (A) was acoustically related to the target (a rhyme). Two other distractors, CS and CA, served as controls for the semantic and acoustic properties of the target, and were generated according to the following scheme used by Coltheart (1977): On one third of the recognition trials, CS was semantically similar to A, and CA was acoustically similar to S; on one third of the trials, CS was semantically similar to A, and CA was acoustically similar to CS; and on the final one third, CS was semantically similar to CA, and CA was acoustically

similar to S.² These relations were used to control for any systematic individual bias that subjects might develop in their responses to distractors. The procedure essentially removes the possibility of subjects' eliminating distractors based on their acoustic or semantic similarity to each other or to the target. These trials were randomly intermixed.

On the recognition test, the five words (T, S, A, CS, CA) on each trial were randomly arranged under the constraint that each type of word appear equally often in each of the five serial positions. The 72 recognition trials were arranged on three sheets of paper, each sheet containing 24 sets. Within the 24 sets on each page there were 8 that contained a target word for each of the three numbers of rehearsals. The order in which the target words appeared in the recognition test was randomized with respect to their order of appearance on the study trials.

Procedure. Subjects were told that the experiment involved a test of short-term memory for numbers. Each trial began with a ready signal given by the experimenter. After 2 seconds, a slide displaying two, three, or four 2-digit numbers appeared for 2 seconds of silent study. It was replaced by a slide presenting a pair of words one next to the other, for 1 second (a short duration was used to prevent covert rehearsal while the words were on the screen). This was followed by a blank slide. The blank slide remained for an interval of 2.7, 8.0, or 13.3 seconds, allowing 1, 5, or 10 overt rehearsals for each pair of words. During the interval, a 75-millisecond, 1000-Hz tone recorded on one track of a tape was sounded at a rate of once every 1.33 seconds. Sub-

² The CS and CA notation should not confuse the reader. It does not refer to the relationship that CS has with S and CA with A. Rather it refers to the fact that CS is semantically related to the items for which it is a control, and CA is likewise acoustically related to the items for which it is a control. We have used this somewhat confusing notation for no reason other than precedence.

TABLE 5
MEAN PERCENTAGE OF CORRECT RESPONSES (WITH
SD IN PARENTHESES) FOR NUMBER RECALL AS A
FUNCTION OF NUMBER OF REHEARSALS AND
PRIMARY TASK LOAD

Primary task load	Number of rehearsals		
	1	5	10
2	85.2 (15.6)	80.7 (14.3)	78.3 (15.8)
3	45.7 (15.4)	35.1 (16.9)	32.4 (18.1)
4	31.4 (14.1)	25.8 (12.9)	22.6 (16.6)

jects were instructed to overtly cycle through both words of each pair in time with the tones, once for each tone sounded. On each trial, after the last tone sounded and the last rehearsal was complete, subjects were given 10 seconds to write the numbers down in their order of appearance.

After completion of all 86 trials, each subject was engaged in conversation for about 2 minutes. Then each subject was unexpectedly given a self-paced recognition test for the words. Subjects were not permitted to change a response to a trial on the recognition test once it was completed.

Following the recognition test, subjects were debriefed and asked whether or not they had anticipated the recognition test. None had.

Subjects were run in groups of three to seven. To avoid interference from overt rehearsals, subjects were seated in individual booths so that they could not see each other. In addition, each subject wore earphones that blocked most external noise.

RESULTS

Number recall. Table 5 presents the percentage of numbers recalled as a function of primary task load and number of rehearsals. In scoring these data, a strict criterion was used in which a response was scored as correct only if all digits were recalled in their correct serial order.

The data in Table 5 show two trends that were confirmed by an analysis of variance. First, recall performance declines as pri-

mary task load increases, $F(2,42) = 6.77$, $p < .01$. Second, recall performance drops with increasing number of rehearsals, $F(2,84) = 4.21$, $p < .05$. Both of these effects are to be expected in a task of this sort. As primary task load increases, there is an increasing capacity load on the memory system, and previous research has shown that this causes a decrease in overall performance (Murdock, 1961). Also, the standard effect of a distractor activity on number recall is that there is interference with the retained items as a direct function of the amount of distractor activity (Peterson & Peterson, 1959).

Word recognition. Table 6 presents accuracy data for the surprise recognition test. A two-way analysis of variance of these data revealed a reliable effect of primary task load, $F(2,42) = 3.97$, $p < .05$, no significant effect of number of rehearsals, $F(2,84) = 1.96$, $p > .05$, and a statistically significant interaction of these variables, $F(4,84) = 4.13$, $p < .01$. To further analyze the interaction, we performed separate one-way ANOVAs for each primary task load. Only that for a task load of 2 showed a reliable effect, $F(2,28) = 3.77$, $p < .05$. Post hoc analysis of this effect revealed that the difference between 1 and 5 rehearsals is reliable, ($t(14) = 3.11$, $p < .05$), but the difference between 5 and 10 is not, ($t(14) = 0.63$, $p > .05$).

Errors in recognition. The design of the experiment permitted an examination of whether subjects tended to make errors on the recognition test by choosing items that

were acoustically or semantically related to the target item. Recall that for each five-alternative recognition item, one of the alternatives (A) was acoustically similar to the target, and that one (CA) was a corresponding control for the acoustically similar alternative (refer to the method section for defining characteristics of the control items). Thus, the ratio A:CA calculated over items for a subject is an index of whether each subject chooses an acoustically similar alternative more than expected by chance when an error is made. Also recall that one of the remaining two alternatives for each recognition item was semantically similar to the target (S), and that the other was a control for semantic similarity (CS). Thus, the ratio S:CS represents a measure of semantic errors committed by each subject.

Averaging these ratios across subjects required first converting each ratio to a log ratio for each subject for each primary task load and for each number of rehearsals. These log ratios were then averaged across subjects, then transformed back to ratios. This method of averaging is necessary because averaging the ratios themselves will lead to a bias. For example, if one subject had a ratio of 2:1 and if this were averaged with another subject whose ratio was 1:2, the result would be 1.25:1. Since a ratio larger than 1.0 leads one to suspect that subjects are making acoustic or semantic errors (depending on which ratio is being computed), this is not a particularly conservative procedure (but see Glenberg & Adams, 1978, for a case where it was used nevertheless). Averaging log ratios is more conservative.

Table 7 presents these averaged ratios for each primary task load and each number of rehearsals. Two-way analyses of variance were computed for the A:CA and the S:CS measures separately. There were no reliable effects in the analysis of the S:CS ratio. By contrast, for the A:CA ratio, all effects were statistically significant: primary task load, $F(2,42) = 3.35$, $p < .05$;

TABLE 6

MEAN PERCENTAGE CORRECT (WITH *SD* IN PARENTHESES) FOR THE UNEXPECTED RECOGNITION TEST AS A FUNCTION OF PRIMARY TASK LOAD AND NUMBER OF REHEARSALS

Primary task load	Number of rehearsals		
	1	5	10
2	33.6 (14.4)	43.1 (16.3)	46.7 (20.7)
3	28.1 (12.6)	33.8 (17.1)	35.8 (16.8)
4	32.5 (12.9)	31.2 (19.1)	27.9 (16.6)

TABLE 7

MEAN RATIOS OF ACOUSTIC AND SEMANTIC CHOICES (WITH *SD* IN PARENTHESES) ON THE RECOGNITION TEST AS A FUNCTION OF PRIMARY TASK LOAD AND NUMBER OF REHEARSALS

Primary task load	Number of rehearsals			
	1	5	10	
A:CA	2	0.98 (1.38)	1.17 (1.29)	1.27 (1.56)
	3	1.10 (1.11)	1.56 (1.53)	2.00 (1.22)
	4	1.03 (1.28)	1.70 (0.99)	1.68 (1.09)
S:CS	2	1.12 (1.49)	1.40 (1.17)	1.30 (1.39)
	3	1.40 (1.13)	1.34 (1.60)	1.48 (1.58)
	4	0.99 (1.50)	0.90 (1.24)	1.23 (1.41)

Note. See the text for details about the items themselves, and for the averaging procedure.

number of rehearsals, $F(2,82) = 3.71, p < .05$; interaction, $F(4,84) = 2.88, p < .05$. For the A:CA measure, separate one-way ANOVAs for each task load revealed reliable effects of number of rehearsals for task loads 3, $F(2,28) = 3.67, p < .05$, and 4, $F(2,28) = 3.67, p < .05$, but not for 2, $F(2,28) = 1.74, p > .05$.

Contingency tabulation and odds ratio analysis. As described above, analysis of the recognition scores revealed a reliable effect of number of rehearsals for a primary task load of 2, and no reliable effects for task loads of 3 and 4. Furthermore the effect for task load 2 was confined largely to the difference between 1 and 5 rehearsals. To assess whether this effect could be attributed to subjects trading resources between memory for the numbers and rehearsal of the words, an analysis was conducted on the recognition and recall scores for the three task loads. This analysis involved calculation of an odds ratio (as described above) for each subject, for each task load, and for each rehearsal interval. Log odds ratios were used to average across subjects. For cases in which any of the cells included a value of 0, .5 was used to replace this value (this is a standard procedure; see Anscombe, 1956; Gart, 1966; Gart & Zweifel, 1967).

The values from this calculation transformed back to odds ratios are shown in Table 8. If the reliable increase in recogni-

TABLE 8

MEAN ODDS RATIOS OF RECALL VS RECOGNITION MEASURES AS A FUNCTION OF NUMBER OF REHEARSALS AND PRIMARY TASK LOAD

Primary task load	Number of rehearsals		
	1	5	10
2	1.09	0.98	0.93
3	1.10	1.34	0.80
4	1.18	1.06	0.70

tion performance between 1 and 5 rehearsals for task load 2 were due to increased trade-off between resources applied to number retention and resources applied to rehearsal, then the odds ratio should show a reliable decrease from 1 to 5. It does not. Nor are the ratios for 1 or 5 rehearsals less than 1.0.

By contrast, there is a decline in odds ratio between 5 and 10 rehearsals for all three primary task loads, as Table 8 illustrates, $F(1,56) = 4.43, p < .05$, and the value for 10 rehearsals is less than 1.0, $t(44) = 2.41, p < .05$. One may therefore conclude that as the rehearsal period lengthens past 5 rehearsals, subjects begin to allocate their processing resources to the act of either rehearsal or retaining the numbers in memory. This trade-off makes it difficult to interpret the lack of increase in recognition scores from 5 to 10 rehearsals. If subjects, for example, are trying to stem erosion of the memory traces for the numbers by concentrating on retaining the numbers, the trade-off indicates that the resources to do this are being borrowed from rehearsal. This, in turn, could lead to lower recognition scores later. This strategy, or others consistent with an increased trade-off between 5 and 10 rehearsals, renders the recognition comparisons between these rehearsal conditions ambiguous. The trade-off for 10 rehearsals does, however, confirm the assumption introduced by Glenberg and Adams (1978) that subjects can share resources between rehearsal and retention of the primary task material. Thus, it is appropriate to assume that dual-task methodology can be used to vary re-

sources applied to the two relevant processes. In particular, we can now feel confident that a demanding primary task load will minimize resources that are allocated to rehearsal.

The odds ratio analysis has revealed two aspects of performance in this task. First, by virtue of the significant drop of the ratio below 1.0 for 10 rehearsals, we can be confident that subjects can trade resources between these tasks and that the odds ratio is sensitive to this strategy. Second, by virtue of the lack of reliable change in ratio between 1 and 5 rehearsals, and by virtue of the fact that these ratios do not differ from 1.0, we can conclude that the change in recognition scores from 1 to 5 rehearsals for task load 2 and the invariance of these scores for task loads 3 and 4 are not due to subjects' trading resources between maintaining the numbers in memory and rehearsing.³

DISCUSSION

The purpose of the present experiment was to assess whether maintenance rehearsal plays any role in laying down memory traces that outlast the duration of the rehearsal process itself. To meet this goal, we refined an experimental procedure devised by Glenberg and Adams (1978) in which maintenance rehearsal was defined by its minimal demand on some central pool of cognitive resources. Let us review the results of our experiment and draw some conclusions about the effect of rehearsal on long-term memory traces, and

³ Given the complexity of the task and of the analyses, a separate experiment was conducted to replicate our findings. The experiment was substantially similar to the present one. This experiment resulted in an effect of task load on recognition scores. Furthermore, the interaction of task load and number of rehearsals showed again, as in the present experiment, that most of the recognition effect was for small numbers of rehearsals with a task load of 2. Finally, these results could not be explained by a trade-off hypothesis, as revealed by an analysis of odds ratios. Our success in replication increases our confidence in the major results of the present experiment.

let us examine some methodological implications of our procedure.

To begin, maintenance rehearsal does seem to play a role in laying down long-term memory traces. This is indicated by three results. First, for a concurrent primary task load of 2, recognition performance increased as subjects increased their rehearsals of the distractor words. Now one might argue, as we did above, that this result is not completely convincing since a concurrent task load of 2 may not sufficiently tax the supply of processing resources to ensure that minimum capacity is devoted to rehearsal (as the current operational definition of maintenance rehearsal demands). The response to this argument is that there are effects of maintenance rehearsal on later recognition with the more demanding concurrent task loads of 3 and 4 as well. But these effects are not increases in accuracy. Rather, subjects make nonrandom patterns of errors in their responses, patterns that indicate residual memory traces of the rehearsed material. For task loads of 3 and 4 the errors were acoustically related to the target. Thus, even though overall recognition accuracy does not increase with more demanding concurrent tasks, there is reason to believe, nonetheless, that the rehearsal process, which with heavier task loads more closely satisfies the minimal capacity definition of "maintenance" rehearsal, is causing more permanent traces to be established.

How do we reconcile these results with the currently popular views of memory? The data contradict the most direct reading of the levels-of-processing view since one of its central assumptions is that maintenance rehearsal keeps a single level of code intact, and does not create the deeper level codes necessary for longer term retention. But suppose this assumption were relaxed, and maintenance rehearsal were assumed to provide an opportunity for deeper level codes to be created, much as the levels-of-processing view assumes for elaborative rehearsal. And further suppose that the ex-

tent to which these deeper level codes are created is a function of the effort or capacity devoted to the maintenance process. This confluence of suppositions provides an interesting opportunity to propose a mechanism that would account for the pattern of errors with changing primary task loads. With little draw on capacity (in the condition where load = 2), as the rehearsal process continues, traces are established with sufficient strength and depth to serve as the foundation for accurate later recognition. As some of the capacity for rehearsal is withdrawn (with loads of 3 or 4), the traces are still encoded to an acoustic level, but not sufficiently well to serve as the basis of accurate recognition, only well enough to cause acoustic confusion errors.

Note that the data suggest that changes in recognition performance (whether accuracy or patterns of errors) are restricted largely to the interval from 2 to 6 rehearsals (data of Geiselman & Bjork, 1980, support this result, although recall data by Rundus, 1980, contradict it). This may suggest that it is only these first few rehearsals that are consequential for the establishment of more permanent memories. Perhaps subjects use the first few rehearsals to somehow integrate the rehearsed words into a unitary trace, a process that may increase the distinctiveness of the words and enhance later recognition (Mandler, 1979). An integration process would result not only in better later recognition, but also, as Mandler has suggested, in a more automated rehearsal process which would free up capacity for number retention in our task. This would be consistent with the trade-off between rehearsal and number retention that we obtained with 10 rehearsals.

While this model of maintenance rehearsal and its effects on memory is somewhat speculative, it is consistent with the available data, and it raises the interesting theoretical possibility that maintenance rehearsal is actually composed of two sub-processes. The first creates traces of the rehearsed items by some operation, such as

integration, that has an effect on later retention. The second takes these traces and recirculates them on a temporary basis to refresh the current contents of memory. This second, more automatic process may have little effect by itself on later retention of the rehearsed material. Taken together, these two processes may form a better view of maintenance rehearsal than the single process account that has been assumed in previous research. We have obtained support for such a two process model in recent work in our laboratory (Naveh-Benjamin & Jonides, 1984).

Our appraisal of the role of rehearsal on later recognition has hinged crucially on expanding the definition of maintenance rehearsal proposed by Glenberg and Adams (1978). They rested their definition on the criterion of capacity devoted to rehearsal, claiming that maintenance rehearsal should be identified with a minimal allocation of processing capacity. There are two implications of this definition that should be made clear.

First, to apply this definition carefully in an experimental context one must be sure that minimal capacity is devoted to rehearsal. Currently, the best available technology to apply to this problem is the dual-task methodology. But there is an important caveat to keep in mind: Effective application of the dual-task technique requires that both tasks draw from the same limited pool of processing resources. This can be tested by examining whether there exist any trade-offs in performance between the two tasks under any of the relevant experimental conditions. In the present experiment, for example, clear evidence of a trade-off was present when subjects were required to produce 10 rehearsals. But evidence for a trade-off is both good news and bad: While it satisfies an assumption of the dual-task methodology, it opens the possibility that differences in later recognition with increased rehearsals are merely a function of the trade-off itself. To rule out this alternative, one must be sure that in-

creased recognition scores are not correlated with increased trade-off between recall and rehearsal.

The second implication of using a capacity-based definition of maintenance rehearsal concerns the relationship between maintenance and elaborative rehearsal. As Glenberg and Adams (1978) discuss, this definitional scheme opens the possibility of viewing maintenance and elaborative rehearsal on a continuum from less to more capacity-demanding processes. Since accumulating evidence now suggests that the two rehearsal processes may not be distinguishable by their differential effects on long-term memory, and since there may be little else to distinguish them formally, the view that they differ from one another only quantitatively becomes more attractive. Perhaps this view will spawn lines of research that compare performance implications of the two rehearsal strategies while at the same time varying the capacity available for each.

REFERENCES

- ANSCOMBE, F. J. (1956). On estimating binomial response relations. *Biometrika*, *43*, 461–464.
- 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.
- COLTHEART, V. (1977). Recognition errors after incidental learning as a function of different levels of processing. *Journal of Experimental Psychology: Human Learning and Memory*, *3*, 437–444.
- 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.
- GART, J. J. (1966). Alternative analyses of contingency tables. *Journal of Royal Statistical Society Series B*, *28*, 164–197.
- GART, J. J., & ZWEIFEL, J. R. (1967). On the bias of various estimators of the logit and its variance, with applications to quantal bioassay. *Biometrika*, *54*, 181–187.
- GEISELMAN, R. E., & BJORK, R. A. (1980). Primary versus secondary rehearsal in imagined voices: Differential effects on recognition. *Cognitive Psychology*, *12*, 188–205.
- GLENBERG, A., & ADAMS, F. (1978). Type I rehearsal and recognition. *Journal of Verbal Learning and Verbal Behavior*, *17*, 455–463.
- 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. (1980). Simpson's paradox and the analysis of memory retrieval. *Psychological Review*, *87*, 398–410.
- JACOBY, L. L., & BARTZ, W. H. (1972). Rehearsal and transfer to LTM. *Journal of Verbal Learning and Verbal Behavior*, *11*, 561–565.
- MAKI, H., & SCHULER, J. (1980). Effects of rehearsal duration and level 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 Experimental Psychology*, *63*, 593–600.
- 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.
- MURDOCK, B. B., JR. (1961). The retention of individual items. *Journal of Experimental Psychology*, *62*, 618–625.
- NAVEH-BENJAMIN, M., & JONIDES, J. (1984). Maintenance rehearsal: A two-component analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*, 369–385.

- NELSON, T. O. (1977). Repetition and depth of processing. *Journal of Verbal Learning and Verbal Behavior*, **16**, 151-171.
- PETERSON, L. R., & PETERSON, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology*, **58**, 193-198.
- ROENKER, D. L. (1974). Role of rehearsal in long-term retention. *Journal of Experimental Psychology*, **103**, 368-371.
- 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**(3), 220-230.
- WINER, B. (1962). *Statistical principles in experimental design*. New York: McGraw-Hill.
- 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 December 30, 1982)

(Revision received June 10, 1983)