Degree of learning, interpolated tests, and rate of forgetting

ROBERT J. ROSE
Memorial University of Newfoundland, Newfoundland, Canada

The purpose of the two experiments reported here was to observe the effects of degree of learning, interpolated tests, and retention interval, primarily on the rate of forgetting of a list of words, and secondarily on hypermnesia for those words. In the first experiment, all the subjects had one study trial on a list of 20 common words, followed by two tests of recall. Half of the subjects had further study and test trials until they had learned the words to a criterion of three correct consecutive recalls. Two days later, half of the subjects under each learning condition returned for four retention tests, and 16 days later, all the subjects returned for four tests. Experiment 2 was similar, except that all the subjects had at least three study trials followed by four recall tests on Day 1, intermediate tests were given 2 or 7 days later, and they all had final tests 14 days later. The results showed that rate of forgetting was attenuated by an additional intermediate set of tests but not by criterion learning. Hypermnesia was generally found over the tests that were given after a retention interval of 2 or more days. The best predictor of the amount of hypermnesia over a set of tests was the difference between overall cumulative recall and net recall on the first test of the set.

An old controversy which has continued to evoke interest over the years concerns the effect of testing on memory performance. The interest has centered in particular on the following question: Does testing attenuate the rate at which memory for a list of stimuli is forgotten, or does it merely elevate memory performance? Slamecka and Katsaiti (1988) addressed this question by arguing that an interpolated test produces some learning in addition to that attained on the study trials. This "overlearning" is latent, in that it does not affect immediate memory performance but it does produce a higher level of performance once forgetting begins to occur. The crucial test of forgetting involves a comparison of the slopes of the memory performance functions for the test and no-test conditions beyond the point where forgetting has begun.

Slamecka and Katsaiti (1988) gave their subjects a test trial after one or more study trials. They found that this additional test enhanced memory performance on a test given 24 h later but the slopes of the functions for the test group and no-test group were equal beyond Day 1. Since Slamecka and Katsaiti found no significant differences in these slopes over three experiments, they concluded that extra testing elevated performance but did not affect the underlying rate of forgetting.

The conclusions of Slamecka and Katsaiti (1988) stand in contrast to those of other studies (e.g., Allen, Mahler, & Estes, 1969; McDaniel & Masson, 1985) that indicated that testing attenuated forgetting and did not merely elevate performance. However, Slamecka and Katsaiti dismissed these studies because they employed only one retention interval in their measures of immediate and delayed memory performance. For them, the necessary and sufficient conditions require measures of performance immediately, after a "moderate" retention interval (i.e., after forgetting has begun), and on at least one other occasion thereafter.

Slamecka and Katsaiti (1988) admit that these conditions were fulfilled in experiments by Runquist (1983, 1986b), who found that tested items were forgotten more slowly than untested items over the succeeding 2 days. (Thereafter, forgetting of the tested items was, if anything, faster than for the untested items, but this interpretation is complicated by floor effects for the untested items under some conditions.) These studies, though, had some other weaknesses. In some experiments, Runquist measured later retention as a function of performance on the prior test. Also, Runquist (1983) encountered some sampling bias among his subjects. For these reasons, Slamecka and Katsaiti considered their results to be stronger evidence than those of Runquist.

The main purpose of the experiments reported here was not to directly challenge the results of Slamecka and Katsaiti (1988) but to test some generalizations of them. A major methodological amendment was the introduction of extra tests following an interval of several days, instead of shortly after study (or between study trials, as in Experiment 3 of Slamecka and Katsaiti, 1988). Bal-

lard (cited in Payne, 1987) suggested that testing had a greater facilitative effect on the retention of successfully recalled material when it occurred at some time after study. If this suggestion is correct, then the extra tests given here could attenuate forgetting in spite of the conclusions of Slamecka and Katsaiti. On the other hand, if the results of Slamecka and Katsaiti showed generalization to the conditions used here, there would be a null effect of interpolated testing on the rate of forgetting.

A second methodological change was the use of free recall of individual words, in order to test the generalizability of the results found with the cued recall of word pairs used by both Runquist (1983, 1986a, 1986b) and Slamecka and Katsaiti (1988). Specifically, in Experiment 1, four groups of subjects were required to learn and free recall a list of words. Two groups (1 and 1T) were given one study trial followed by two recall trials, while two other groups (C and CT) learned the list to a criterion of three consecutive correct recalls. Groups 1T and CT returned 2 days later for four recall trials. Sixteen days after the original learning, all the subjects returned for four recall trials.

Although this paradigm measured memory performance three times, one might argue that a proper test of the learning hypothesis of Slamecka and Katsaiti (1988) would demand *four* times of measurement under the conditions of this experiment. A critical point in their argument is that two of the measures occur after the introduction of the extra tests in order to accommodate the hypothesized latent effects of overlearning. However, the use here of multiple interpolated tests after 2 days rendered the fourth measurement unnecessary for the following reason. The amount of learning induced in groups like 1T and CT by interpolated Test 1 would, to a large extent, be reflected in those groups' performance on Test 2, the learning from Test 2 would be reflected on Test 3, and so on. Of course, there would be learning on Test 4, the last test of the set, which would affect performance on a hypothetical Test 5, but the amount of additional learning from Test 4 onward would probably be very small and could, in any case, be estimated from the trend shown in Tests 1-4. (Underwood, 1964, illustrated this estimation under somewhat similar circumstances.) The point is, if additional testing had no further effect after four trials, then the learning argument of Slamecka and Katsaiti was accommodated and one could compare the forgetting functions of test and no-test conditions across one retention interval.

Another purpose of this study was to compare rates of forgetting of the same list following near minimal learning for Groups 1 and 1T (one study trial followed after 1 min by two immediate recall tests) with a high level of overlearning for Groups C and CT (three consecutive correct recalls). Slamecka and McElree (1983) stated that "apparently there are no published direct comparisons of forgetting functions for overlearned versus underlearned lists" (p. 394), but went on to imply that the probability of overlearning affecting the rate of forgetting was small.

One should note that Group CT was expected to show less forgetting over the first 2 days than Group 1T because CT's "overlearning" of the list, like that of the subjects of Slamecka and Katsaiti (1988) who were given extra tests, should produce latent learning which would have no effect on immediate performance. However, the arguments of both Slamecka and McElree and Slamecka and Katsaiti predicted that rate of forgetting after the first couple of days would be the same for groups receiving criterion learning and groups receiving one study trial (excluding the effects of the interpolated tests).

A third aim of this study was to take advantage of the use of multiple tests, which are a critical feature of the procedure, to examine the phenomenon of hypermnesia, which is a net increase in the number of items recalled over a set of test trials (or over a period of time) without the benefit of further study trials. In his review of studies of hypermnesia, Payne (1987) argued that the best predictor of hypermnesia, at least when it occurs with verbal material, is the level of cumulative recall, a view expressed earlier by Roediger, Payne, Gillespie, and Lean (1982). This hypothesis states that those "study conditions that produce high performance levels tend to produce hypermnesia, whereas those that produce lower levels of recall are less likely to obtain it" (Payne, 1987, p. 12). One aim in the present study was to test this hypothesis, in part by examining the correlations between hypermnesia, on the one hand, and cumulative recall, Test 1 recall, and the difference between these two measures, on the other (see Payne, 1986, Table 3).

Based on the cumulative recall hypothesis, three predictions were made. First, there would be a significant positive correlation between the magnitude of hypermnesia and cumulative recall. In addition, the two groups engaging in criterion learning would produce more hypermnesia than the two groups given one-trial learning. Finally, Groups 1 and 1T would produce less hypermnesia at the long retention intervals than they would after a few minutes, because their cumulative recall levels would almost certainly be lower after 2 or 16 days.

As a final point, Rose and Howe (1989) found significant amounts of hypermnesia over four retention trials given 2 weeks after learning a list of words to criterion, but they did not obtain a significant amount of hypermnesia over two tests that followed 1 min after one study trial. Another purpose of Experiment 1, then, was to replicate these results by giving all the subjects two recall tests after one study trial on Day 1 (as well as four retention tests 16 days later and, for Groups 1T and CT, four recall tests after 2 days).

To recapitulate, the purpose of the present experiment was to determine if: (1) rate of forgetting was affected by a series of free-recall tests interpolated after 2 days between initial learning and final free-recall tests, (2) rate of forgetting was affected by the overlearning of a list (relative to near-minimal learning of the list), either after the first 2 days or over the last 14 days of the retention

interval or both, and (3) hypermnesia was present on the sets of retention tests and was predictable from the level of cumulative recall.

EXPERIMENT 1

Method

Subjects. The subjects were 80 undergraduates of Memorial University who were randomly assigned to one of four groups, 1, 1T, C, or CT. Subjects in Groups 1 and 1T had one learning trial while those in Groups C and CT learned to a criterion. Subjects in Groups 1T and CT also had intermediate tests 2 days after learning. The subjects were volunteers and were paid for their participation.

Procedure. On Day 1, each subject was presented with a list of 20 common words (A or AA in the Thorndike-Lorge word list), one word at a time, on a video display terminal. Each word was exposed for 3 sec; the order of presentation was random, as determined by a programmed computer. Following the initial presentation of the list, the subject engaged in a maze puzzle distractor task for 1 min and then was required to free recall the words on the list for 3 min. This test was followed by another minute on the distractor task and another 3 min of recall. After the initial recall tests, the subjects in Groups 1 and 1T were dismissed. The subjects in Groups C and CT continued to have sequences of learning trial, distractor task, and test trial until they achieved a criterion of three consecutive correct recalls. (The words appeared in a different random order on each learning trial for each subject.)

The subjects in Groups 1 and C were asked to return in 16 days (i.e., on Day 17) and those in Groups 1T and CT were asked to return in 2 days (on Day 3). No subject was informed of the reason for the return visits. During these second sessions, each subject was given four free-recall test trials of 3 min each, with 1 min of the maze distractor task intervening between tests. In addition, the subjects in Groups 1T and CT were asked to return for a third session on Day 17, that is, 16 days after the original learning, for a further set of four free-recall tests (with intervening distractor task).

Finally, at the conclusion of the last test session, each subject was asked two questions: (1) Did you expect to be tested for recall of the words during the return sessions? (2) Did you rehearse the words during the retention interval(s)?

Results and Discussion

Learning-session performance. An analysis of variance of the number of words recalled on the initial retention tests (hereafter called 1a and 1b) was carried out with group and test as main factors. This analysis showed a significant difference between Tests 1a and 1b, with means of 9.51 and 9.23 for 1a and 1b, respectively [$F(1,76) = 4.23$, $MS_e = 0.78$]. There were no significant differences among the four groups and there was no significant group × test interaction [$Fs(3,76) < 1$]. (Note that all reported effects are significant with $p < .05$, unless otherwise noted.) The next step in the analyses of the data was to test whether there were any differences between Groups C and CT in the attainment of the criterion. The mean number of trials and the mean number of errors (of omission and commission) to achieve criterion were 8.35 and 23.95, respectively, with no differences between Groups C and CT on either measure [$ts(38) < 1$].

These analyses of the learning data indicate that the random assignment of subjects achieved its goal by producing groups that did not differ in performance, at least after one learning trial and, for Groups C and CT, in the attainment of the criterion. There is, therefore, no reason to assume that Group 1, if they had a test of recall after 2 days, would differ significantly from Group 1T on that test. Similarly, the hypothetical results for Group C after 2 days would not differ significantly from those for Group CT. These hypothetical results are incorporated into retention curves 1 and C in Figure 1. These curves depict the forgetting functions for groups receiving one learning trial and criterion learning, respectively, without the effects of intermediate tests. The single initial datum for Groups 1 and 1T represents their mean recall on Test 1b.

Rate of forgetting. There are two equivalent ways of determining rate of forgetting: the absolute number of
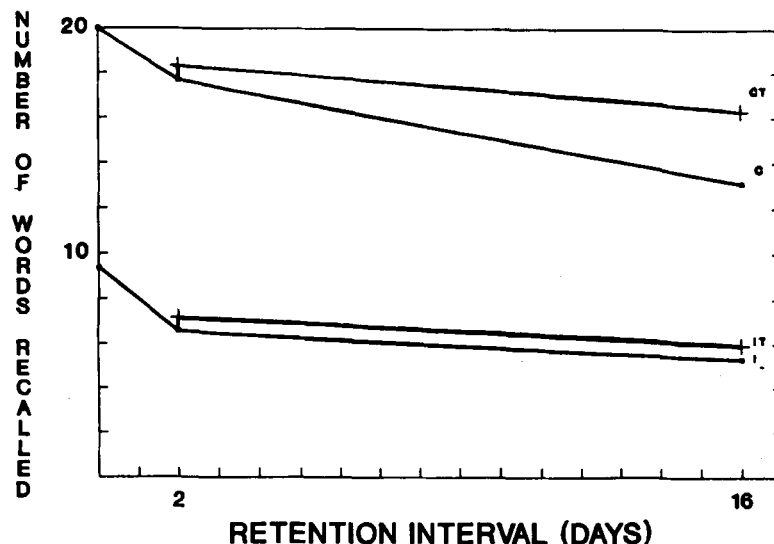

Figure 1. Number of words recalled by retention interval and condition in Experiment 1.

words forgotten over equal retention intervals, or the slope of the forgetting function over those intervals (see Slamecka & McElree, 1983). Given the nature of the present experiment, in which all the subjects were tested on Day 1 and 16 days later but only some subjects were given intermediate tests, each of these measures was used as determined by the situation. Since retention interval was neither a completely within- nor completely between-subjects variable, an omnibus analysis of variance across all intervals could not be done. Instead, separate analyses for the first 2 days and the last 2 weeks were performed.

First, the mean number of words forgotten by Group 1T (2.50) and Group CT (2.35) between the last test on Day 1 and the first retention test 2 days later was found not to differ significantly [$t(38) < 1$]. Contrary to the prediction based on the results of Slamecka and Katsaiti (1988), the amount of forgetting over 2 days for the "overlearning" Group CT was not less than the forgetting by Group 1T given only one study trial.

Next, for the period between Day 3 and Day 17 (i.e., 2 and 16 days, respectively, after Day 1), recall performance on the first tests of Day 3 (Groups 1T and CT) and of Day 17 (Groups 1 and C) was analyzed with level of learning and retention interval as between-subject factors. These data represented the forgetting function over the last 14 days without the effects of the multiple interpolated tests. The important result from this analysis was the presence of an interval × level interaction of borderline significance [$F(1,76) = 4.09$, $MS_e = 13.31$, $p = .047$]. This result indicates that, contrary to expectations, the rate of forgetting over the last 14 days of the retention interval was affected by the level of learning. This outcome will be addressed later.

The effects of the interpolated tests on forgetting can be seen in curves 1T and CT in Figure 1. When one turns to an analysis of these effects, a problem arises. An analysis of the recall performance after 16 days with interpolated tests as a factor is not appropriate, because the interpolated tests could elevate the performances of Groups 1T and CT (e.g., due to hypermnesia) without affecting the rate of forgetting following the last test. What is required is a comparison of the number of words forgotten over the last 14 days by the four groups. This was done for each subject in Groups 1T and CT by subtracting the result on Test 1, Day 17, from the result on Test 4, Day 3. However, it is literally impossible to do this for Groups 1 and C, because they had no tests on Day 3. The best one can do is *estimate* the amount of forgetting for Groups 1 and C over the last 14 days by using the mean recall performances of Group 1T (6.55) and Group CT (17.65) on Test 1 of Day 3 as the means on Day 3 of Groups 1 and C, respectively. As discussed previously, there is no reason to assume that Groups 1 and C, *if* they had a test after 2 days, would differ from Groups 1T and CT, respectively. Accordingly, the forgetting rate for each subject in Groups 1 and C from Day 3 to Day 17 was estimated by subtracting the result

on Test 1, Day 17 from 6.55 for Group 1 and from 17.65 for Group C.

The analysis of these forgetting data showed that Groups C and CT, who learned to a criterion, had a greater loss of retention over the last 14 days than did Groups 1 and 1T [$F(1,76) = 6.15$, $MS_e = 13.67$]. A comparison of individual means showed that this result was due to Group C's forgetting ($M = 4.55$) significantly exceeding the amount forgotten by the other three groups ($M = 1.25$, 1.20, and 2.00 for Groups 1, 1T, and CT, respectively), which did not differ among themselves.

These results indicate that, insofar as Group C exceeded Group CT in number of words forgotten between Day 3 and Day 17, intervening tests did attenuate the rate of forgetting. Also, insofar as Groups 1T and CT had equal amounts of forgetting over the two periods of retention, rate of forgetting of a list is independent of level of original learning. Unfortunately, the number of words forgotten by Group 1 presents a problem for the generality of these inferences because they did not differ from those for Group 1T and were not equal to those for Group C. These unexpected results for Group 1 were also responsible for the significant interaction reported above. The probable reason for these results is that Group 1 was near the "floor" after an interval of 2 days. A t test showed that forgetting for Group 1 did not differ significantly from zero over the following 14 days [$t(19) = 1.87$, $p = .08$]. Otherwise, the rate of forgetting for Group 1 might have approximately equaled the rate for Group C and been significantly greater than the rate for Group 1T. This hypothesis was tested in Experiment 2.

Effects of expectations. As mentioned previously, the subjects were asked at the conclusion of the experiment whether they had expected a retention test on Day 3 and/or on Day 17 and whether they had rehearsed during the retention intervals. Only 8 subjects reported any rehearsal, 5 of whom came from Group 1T and reported some rehearsal following the interpolated tests on Day 3. The number of subjects who expected a test of retention varied considerably, from 1 for Group C to 14 for Group 1T. Again, most of these "non naive" subjects formed their expectations following the interpolated tests.

Analyses of the data for only the naive subjects, who neither expected the tests beyond Day 1 nor rehearsed, showed the same general trends as the analyses reported above. After a retention interval of 2 days, there were 18 naive subjects each, in Group 1T and CT, whose mean number of words forgotten (2.44 and 2.56, respectively) did not differ [$t(34) < 1$]. The between-subjects analysis of recall performances on Test 1 over the last 14 days was similar to that for all the subjects, except that the retention interval × level of learning interaction was not significant. On the third analysis with interpolated tests as a factor, the amount forgotten by Group C (4.65) again was greater than the amount forgotten by the other three groups (1.59, 2.00, and 2.43, for Groups 1, 1T, and CT, respectively), although level of learning did not quite reach

statistical significance this time [$F(1,51) = 3.59$, $MS_e = 18.08$, $p = .064$]. In general, it appears that expectations and rehearsal over the retention intervals had little effect on the results of this experiment. This null effect replicates similar results by Runquist (1983, Experiment 3).

Hypermnesia scores. It has been reported that recall on Test 1a was significantly higher than recall on 1b, following the initial study trial on Day 1. These results obviously reflected forgetting, not hypermnesia, and replicated the results of Rose and Howe (1989). Furthermore, the group × trial interaction was nowhere near statistical significance, so this forgetting trend held for all four groups.

However, after retention intervals of 2 and 16 days, there was significant hypermnesia, as can be seen from the net recall data in Table 1. (The amounts of hypermnesia after an interval of 2 days can also be seen in Figure 1, by comparing curve C with curve CT and curve 1 with curve 1T.) The suffixes "-2" and "-16" in Table 1 denote the retention intervals for the learning conditions. Separate analyses of variance for the net recall data showed that performance increased over the four retention tests on both Day 3 [$F(3,114) = 10.71$, $MS_e = 0.29$ (Groups 1T and CT only)] and Day 17 [$F(3,228) = 24.67$, $MS_e = 0.59$ (all four groups)]. Statistical comparisons of the individual means showed that, for each day, each mean differed from every other mean except for those of Tests 3 and 4.

Two analyses of the amount of hypermnesia, measured as the difference between Test 4 recall and Test 1 recall, were then carried out. The first analysis concerned the hypermnesia found on the initial set of long-term memory tests (Conditions 1T-2, CT-2, 1-16, and C-16), with level of learning and retention interval as factors. This analysis showed that hypermnesia was affected only by

retention interval and then only with borderline significance [$F(1,76) = 3.42$, $MS_e = 1.93$, $p = .068$]. The second analysis, which concerned the hypermnesia after 16 days (Conditions 1T-16, CT-16, 1-16, and C-16), showed that the results were not affected by interpolated tests [$F(1,76) = 2.06$], but were affected by level of learning [$F(1,76) = 3.57$, $MS_e = 2.19$], with borderline significance ($p = .063$).

Since hypermnesia is the positive difference over a series of tests between the number of items recovered on a later test but not on an earlier test (sometimes called reminiscence) and the number of items forgotten between those tests, the data for these two components were examined after the fashion of Payne (1986). The mean number of items recovered (NC) or forgotten (CN) between tests over the four recall trials after intervals of 2 days (Conditions 1T-2 and CT-2) and 16 days (Conditions 1T-16, CT-16, 1-16, and C-16) are shown in Table 2. (The components of the recall data for Tests 1a and 1b of Day 1 were also tabulated, although they do not appear in Table 2. An analysis of the data for Tests 1a and 1b merely confirmed what was reported above, that is, that intertest forgetting exceeded recovery for all four groups.)

An analysis of the recovery data for Conditions 1T-2, CT-2, 1-16, and C-16 with retention interval, level of learning, and serial position of test as factors showed that recovery was greater after the longer interval [$F(1,76) = 5.42$, $MS_e = 0.69$] and after criterion learning [$F(1,76) = 4.07$]. In addition, recovery decreased over the four tests [$F(2,152) = 3.50$, $MS_e = 0.48$]. Intertest forgetting was affected only by level of learning, with the criterion groups producing more forgetting with borderline significance [$F(1,76) = 4.01$, $MS_e = 0.15$, $p = .049$].

Next, the components of hypermnesia for all groups after a retention interval of 16 days were analyzed, primar-

Table 1
Mean Net Recall and Cumulative Recall

Condition	Net Recall				Cumulative Recall	
	Test 1	Test 2	Test 3	Test 4	C_sR	C_wR
Experiment 1						
1T-2	6.55	6.85	7.10	7.15	7.20	
CT-2	17.65	18.00	18.20	18.30	18.50	
1T-16	5.95	5.95	6.30	6.40	6.45	10.35
CT-16	16.30	16.85	17.05	17.30	17.45	20
1-16	5.30	5.65	6.05	6.15	6.30	10.80
C-16	13.10	13.85	14.35	14.60	14.90	20
Experiment 2						
3T2-2	13.60	13.67	13.60	13.67	13.87	
CT2-2	18.47	19.13	19.33	19.53	19.60	
3T7-7	10.33	10.67	10.60	11.00	11.27	
CT7-7	17.13	17.93	18.33	18.73	18.73	
3T2-14	13.07	13.07	13.13	13.20	13.33	15.53
CT2-14	17.47	18.07	18.13	18.27	18.43	20
3T7-14	10.47	10.87	11.13	11.40	11.47	15.37
CT7-14	17.73	17.80	18.00	17.93	18.13	20
3-14	7.93	8.27	8.40	8.53	8.80	15.13
C-14	12.27	12.67	13.13	13.73	14.07	20

Note—Maximum = 20. C_sR = Cumulative recall over a set of four tests. C_wR = Cumulative recall across the whole experiment.

Table 2
Mean Number of Words Recovered (NC) or Forgotten (CN)
Between Tests

Condition	NC			CN		
	Tests 1-2	Tests 2-3	Tests 3-4	Tests 1-2	Tests 2-3	Tests 3-4
			Experiment 1			
1T-2	.35	.30	.10	.05	.05	.05
CT-2	.50	.35	.20	.15	.15	.10
1T-16	.20	.35	.10	.20	.00	.00
CT-16	.70	.35	.35	.15	.15	.10
1-16	.40	.45	.30	.05	.05	.20
C-16	1.00	.65	.50	.25	.15	.25
M	.53	.41	.26	.14	.09	.12
			Experiment 2			
3T2-2	.20	.13	.13	.13	.20	.07
CT2-2	.67	.47	.27	.00	.27	.07
3T7-7	.53	.33	.47	.20	.40	.07
CT7-7	.93	.67	.40	.13	.27	.00
3T2-14	.07	.13	.13	.07	.07	.07
CT2-14	.73	.27	.20	.13	.20	.07
3T7-14	.60	.40	.33	.20	.13	.07
CT7-14	.40	.27	.07	.33	.07	.13
3-14	.47	.33	.20	.13	.20	.07
C-14	.87	.67	.73	.47	.20	.13
M	.54	.37	.29	.18	.21	.08

ily to determine the effects of the interpolated tests. No factor or interaction was significant in the analysis of the forgetting data. The analysis of the recovery data, however, showed that recovery again increased with criterion learning [$F(1,76) = 7.49$, $MS_e = 0.76$] and decreased with interpolated tests [$F(1,76) = 3.99$] and serial position of test [$F(2,152) = 2.92$, $MS_e = 0.52$]. These last two effects were of borderline significance ($ps = .049$ and .057, respectively).

Hypermnesia correlations. Correlations were calculated for hypermnesia with Test 1 recall, cumulative recall, and the difference between cumulative recall and Test 1 recall. These correlations are given in Table 3 for each condition. Note that in this experiment, unlike Payne (1986), there are two measures of cumulative recall. The typical measure is the number of unique words recalled by a subject on a given occasion across a set of tests, shown as C_sR in Table 1. However, given that subjects here were tested on more than one occasion, another measure of cumulative recall is the number of unique words recalled over the entire experiment (C_wR in Table 1). Correlations involving both of these measures appear in Table 3.

Perhaps the most noticeable feature of Table 3 is the set of coefficients of correlation with $C_sR - T1$, the difference between cumulative recall across tests and Test 1. These highly significant positive results are similar to those of Payne (1986). However, they are trivial because, for most subjects, cumulative recall across a set of four tests was equal to Test 4 recall. (Table 1 shows that mean net recalls on Test 4 and mean cumulative recalls are very

similar.) So this particular correlation almost consisted of comparing hypermnesia with itself.

Another feature of these data is that the coefficients of hypermnesia with Test 1 recall and with cumulative recall (calculated across a set of tests) were dissimilar for the one-trial and criterion conditions. The results for the one-trial condition were similar to those found by Payne (1986), that is, near-zero coefficients with Test 1 recall and moderately positive coefficients with cumulative recall. For the criterion learning condition, the coefficients were moderately negative, at least for Test 1 recall. This outcome is not surprising, though, because a high Test 1 recall leaves little room for improvement over succeeding trials when the absolute "ceiling" is fixed. There was also a moderate coefficient of correlation between hypermnesia and C_wR, the cumulative recall calculated across the whole experiment. (Note that this coefficient cannot be calculated for the criterion condition, whose cumulative recall across the whole experiment is fixed at 20. For the same reason, the coefficient for the criterion condition with $C_wR - T1$ is equal in size but opposite in direction to the coefficient with T1 recall alone.)

Statistical treatment of these correlational data is somewhat problematic; because there are relatively few data points contributing to each condition, few of the individual correlations aside from $C_sR - T1$ are significant. One way around this problem is to treat each condition as an independent test of the same hypothesis and employ the statistic $\chi^2 = 2\Sigma - \ln P_i$, where P_i is the probability that the correlation is due to chance. (See Harrison, 1979, for an example of this statistic as applied to a set of correla-

Table 3
Correlations Between Hypermnesia and Test 1 Recall, Cumulative
Recall, and the Difference Between Cumulative Recall and Test 1

	T1	C_sR	C_sR-T1	C_wR	C_wR-T1
			Condition		
			Experiment 1		
1T-2	−.07	.34	.98†	.55*	.73†
1T-16	−.06	.38	.98†	.16	.25
1-16	−.03	.45*	.96†	.36	.43
M	−.05	.39*	.97†	.36*	.46†
CT-2	−.37	−.14	.87†		.37
CT-16	−.68†	−.47*	.93†		.68†
C-16	−.37	−.11	.91†		.37
M	−.47†	−.24	.90†		.47†
			Experiment 2		
3T2-2	−.02	.04	.59*	.02	.11
3T2-14	−.14	−.02	.87†	.03	.33
3T7-7	−.23	.03	.95†	.19	.53*
3T7-14	.05	.27	.96†	.30	.32
3-14	−.27	−.03	.87†	.24	.44
M	−.12	.06	.85†	.16	.35
CT2-2	−.81†	−.02	.99†		.81†
CT2-14	−.49	−.06	.90†		.49
CT7-7	−.78†	.20	1.00†		.78†
CT7-14	−.46	−.21	.79†		.46
C-14	.10	.45	.98†		−.10
M	−.49†	.07	.93†		.49†

Note—T1 = Number of words recalled on Test 1; C_sR = Cumulative recall over a set of four tests; C_sR-T1 = Difference between C_sR and T1; C_wR = Cumulative recall across the whole experiment; C_wR-T1 = Difference between C_wR and T1.   *$p < .05$.   †$p < .01$.

tional measures.) The statistic was used here, not with the six conditions as a set, but rather with the three one-trial and three criterion conditions separately, because, as mentioned above, the trends for these learning conditions were somewhat different on some measures.

The correlation between hypermnesia and C_sR-T1 had, of course, a strong overall positive effect for both the one-trial and criterion conditions [$\chi^2(6) > 42$]. For the one-trial conditions, the correlations between hypermnesia and C_sR, C_wR, and C_wR-T1 all had significant overall effects [$\chi^2(6) > 14.48$]. For the criterion conditions, the correlations for C_wR-T1 and its converse Test 1 were significant overall [$\chi^2(6) = 22.72$]. In summary, the only measures to produce a significant overall correlational effect with hypermnesia, consistent across both learning conditions, were the differences between Test 1 and cumulative recall, calculated either over a set of four tests or across the whole experiment.

Finally, one might argue that a valid correlational relationship with hypermnesia should hold, not only within a condition, but also across conditions. Accordingly, correlations were calculated between hypermnesia and the five measures of Table 3 collapsed across all six conditions. These yielded correlations of −.09, .12, .94, .34, and .30, for measures T1, C_sR, C_sR-T1, C_wR, and C_wR-T1, respectively. The last three were statistically significant (ps < .01). The pattern here, then, was similar to the pattern for the overall effects of the "within" conditions. Only the correlations for the differences be-

tween Test 1 and cumulative recall have consistent effects over all of the conditions. (Recall that the correlation for C_wR cannot be calculated for criterion conditions.)

Hypermnesia summary. The hypermnesia data discussed here contain several points of interest. First, there was net forgetting but no hypermnesia for Tests 1a and 1b on Day 1. However, there was significant hypermnesia across tests for all groups on Days 3 and 17. Second, retention interval and interpolated tests had small, rather inconsistent effects on hypermnesia. Where they had effects, they occurred mainly on recovery rather than on intertest forgetting. Third, there were tendencies for recovery to decline over tests and increase over retention interval. Fourth, the most consistent, nontrivial relationship with hypermnesia was the positive correlation for the difference between Test 1 recall and cumulative recall across the whole experiment.

The tendency for intertrial recovery to decrease over the four test trials while intertrial forgetting remained relatively constant stands in contrast to Payne (1986), who found that forgetting decreased over test trials while recovery was steady or decreased to only a small extent. The net result is that hypermnesia declined over the four tests used here but increased or remained steady over the three tests used (in most cases) by Payne (1986). The decline in hypermnesia over the present test trials is critically important to the arguments made in the introduction regarding the rationale for using only one retention interval after the interpolation of the extra tests on Day 3. Since the

difference between Tests 3 and 4 was not significant, the hypermnesic effect seems to have reached its asymptote. Therefore, any difference between the expected level of performance on a fifth test trial and the Test 4 performance (with the concomitant error in forgetting over the following retention interval) would only be slight.

EXPERIMENT 2

Experiment 1 can be described as only partially successful in rejecting the null hypothesis that additional testing has no effect on rate of forgetting of a list of words. One cannot unequivocally reach a conclusion based on a comparison of the slopes of the test and no-test conditions when the latter show no significant forgetting. Group 1 showed no significant drop in retention between Days 3 and 17. Thus, while the differences between Groups C and CT indicate that interpolated tests attenuate forgetting, the differences between Groups 1 and 1T are uninterpretable. It was suggested above that Groups 1 and 1T did not differ in rate of forgetting as a result of a floor effect. To test this hypothesis, three groups of subjects in Experiment 2 were given three learning trials, instead of the single learning trial used in the previous experiment. The assumption was made that three trials would induce an initial level of performance such that there would be a significant difference in retention after 2 weeks between those subjects given intermediate tests and those given none. A second amendment was the use of two intermediate retention intervals (2 days and 7 days) and not just one. The second intermediate retention interval allowed for an additional test of the rates of forgetting under the experimental conditions.

Another aim of Experiment 2 was to explore further the presence of hypermnesia. In Experiment 1, significant hypermnesia was found after retention intervals of both 2 and 16 days, but not after a few minutes. The absence of hypermnesia on Day 1 may have been due to the effects of recency. Hypermnesia will occur only if there is some difference between what is available in memory and what is currently retrievable. Unless some forgetting has occurred to produce this difference, there will be no hypermnesia. To test this argument, all the subjects were given four tests of retention after the first three learning trials on Day 1 (as well as four tests of retention at one or more later dates). If the results of Experiment 1 were replicated, recall on Test 1b of Day 1 would be less than recall on Test 1a. However, if the preceding argument held, recall following this initial forgetting would increase over the next two trials (Tests 1c and 1d).

Method

Subjects. The subjects were 90 undergraduate volunteers of Memorial University who were randomly assigned to one of six groups, 3, 3T2, 3T7, C, CT2, or CT7. Subjects in Groups 3, 3T2, and 3T7 had three learning trials, and those in Groups C, CT2, and CT7 learned to the criterion. In addition, subjects in Groups 3T2 and CT2 had interpolated tests 2 days after learning, while the subjects in Groups 3T7 and CT7 had extra tests 7 days after learning. The subjects were paid for their participation.

Procedure. The procedure was the same as in Experiment 1, except for the following changes. On Day 1, all the subjects received three study trials on the list of 20 words followed by four test trials. The subjects in Groups 3, 3T2, and 3T7 were then dismissed, while the subjects in Groups C, CT2, and CT7 continued until they had learned the list to the criterion, as in Experiment 1. All the subjects were asked to return in 2 weeks for a further unspecified task, which, in fact, was a set of four recall tests with 1 min of maze puzzles intervening between the tests. In addition, Groups 3T2 and CT2 returned after 2 days and Groups 3T7 and CT7 returned after 7 days for a set of four recall tests (again with no explicit warning and with intervening maze puzzles).

Results and Discussion

Learning-session performance. The means across all six groups for the four tests that followed the three learning trials on Day 1 were 14.57, 14.22, 14.39, and 14.49, respectively. An analysis of the data showed that recall on Test 1b was significantly below that for Tests 1a and 1d [$F(3,252) = 3.57$, $MS_e = 0.56$], a result that supports the argument made in the introduction to this experiment. Neither group nor the group $\times$ test interaction was near significance as a factor ($Fs < 1.12$). The number of trials and number of errors for Groups C, CT2, and CT7 to reach criterion were also analyzed; they did not differ significantly on either measure [$Fs(2,42) < 1.27$]. The means were 8.04 and 21.58 for trials and errors, respectively.

Rate of forgetting. The retention data for the six groups are shown in Figure 2. Following the convention of Experiment 1, the initial datum for Groups 3T2, 3T7, and 3 is their mean for the last test on Day 1, and the data for the groups given no intermediate tests are hypothetical for the intermediate days. Hence, curves C and 3 show, respectively, the recall performance on the first interpolated test for Groups CT2 and CT7 and Groups 3T2 and 3T7. The final part of curves CT2, CT7, 3T2, and 3T7 show recall for these groups on the last interpolated test and Test 1 on the final day. (As in Experiment 1, the first parts of the forgetting curves for these groups are superimposed over curves C or 3.)

The data were analyzed in the same manner as in Experiment 1, and for the same reasons. First, an analysis of the number of words forgotten after an interval of 2 days (by Groups 3T2 and CT2) showed that Group 3T2, with a mean of 0.20, forgot significantly fewer words than did Group CT2, with a mean of 1.53 [$t(28) = 2.24$, $SE = 0.69$]. This result failed to replicate the forgetting data for the 2-day retention interval in Experiment 1 but it also was directly contrary to the prediction based on Slamecka and Katsatai (1988). The very low number of words forgotten by Group 3T2 is probably an anomaly. In fact, over the entire retention interval of 14 days, Group 3T2 forgot an average of only 0.73 words, a decline that did not differ significantly from zero [$t(14) = 1.11$].

Next, the number of words recalled on Test 1 after retention intervals of 2 days (Groups 3T2 and CT2), 7 days (Groups 3T7 and CT7), or 14 days (Groups 3 and C) was analyzed with level of learning and retention interval as between-subject variables. Both main effects were, of course, significant but the important result was that the level
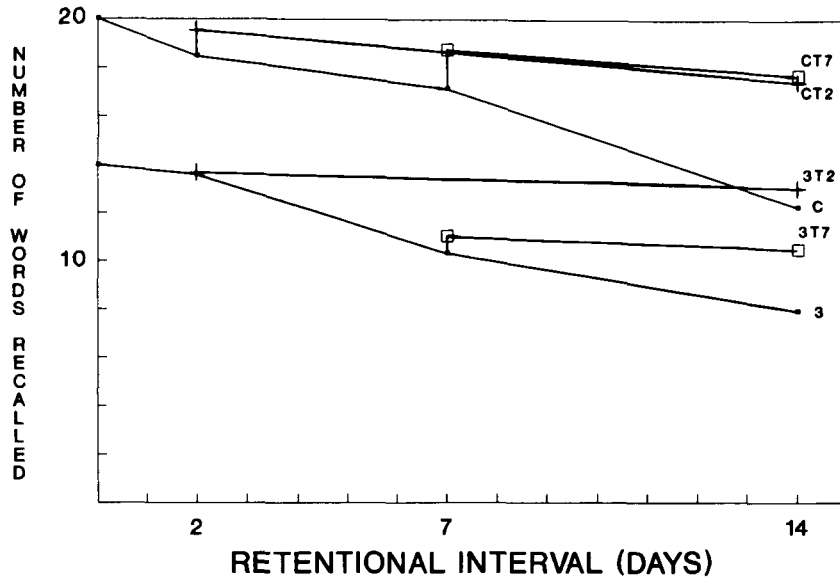
Figure 2. Number of words recalled by retention interval and condition in Experiment 2.

X interval interaction was nowhere near significant [$F(2,84) = 1.00$]. This result supported the conclusion of Slamecka and McElree (1983) that rate of forgetting of a list of words is independent of the level of learning of that list. It also supported the contention that the unexpected interaction found in the corresponding analysis of Experiment 1 resulted from a floor effect on Group 1.

Separate analyses of forgetting over the last 12 days by Groups 3T2, CT2, 3, and C and over the last 7 days by Groups 3T7, CT7, 3, and C were carried out. As in Experiment 1, the data points for Groups 3 and C were estimates of forgetting made by subtracting the recall scores on Test 1 of the final day from the means for Test 1 on Day 3 or Test 1 on Day 7 for 3T2, CT2, 3T7, and CT7, respectively. These analyses showed that the interpolated tests had a significant effect over both the last 12 days of retention [$F(1,56) = 36.13, MS_e = 8.77$] and the last 7 days [$F(1,56) = 12.81, MS_e = 8.94$]. Neither level of learning nor its interaction with interpolated tests had a significant effect for either retention interval.

These results replicated those of Experiment 1 in that rate of forgetting was not affected by overlearning but was considerably slowed by interpolated testing. In fact, after the interpolated tests, only Group CT2 showed an amount of forgetting significantly different from zero [$t(14) = 2.09, SE = 0.51$], although the forgetting by Group CT7 was of borderline significance [$t(14) = 2.09, SE = 0.48, p = .055$]. In addition, these results supported the view that a similar outcome was prevented in Experiment 1 by the constraint of a floor effect on Group 1.

Effects of expectations. The numbers of subjects who expected a test of memory after Day 1 and/or rehearsed during the retention interval were 6, 7, 6, 6, 7, and 10 for Groups 3, 3T2, 3T7, C, CT2, and CT7, respectively. As in Experiment 1, few of these subjects reported actual rehearsal.

In this experiment, the analyses of the data for only the naive subjects showed the same general trends as did the analyses of all of the subjects, with two exceptions. First, the difference in forgetting over 2 days between Group 3T2 (0.75) and Group CT2 (1.63) was no longer significant [$t(14) = 1.21, SE = 0.72$]. This result replicated the findings in Experiment 1. Second, over the last 7 days, there was significantly more forgetting by the criterion Groups C and CT7 (4.46 words) than by Groups 3 and 3T7 (1.89 words) [$F(1,28) = 7.25, MS_e = 7.17$], although the level of learning X interpolated tests interaction was again nonsignificant [$F(1,28) < 1$]. This second finding may be somewhat embarrassing for Slamecka and McElree (1983), considering their conclusion regarding the null effects of level of learning on rate of forgetting. However, this particular result can probably be dismissed as an anomaly arising from relatively few subjects. The larger analysis of number of words recalled by the naive subjects on Test 1 after 2 days (Groups 3T2 and CT2), 7 days (Groups 3T7 and CT7), and 14 days (Groups 3 and C) replicated the aforementioned analysis of all of the subjects in finding the level X interval interaction to be nowhere near significant [$F(2,42) < 1$]. The most important result of the naive subjects, for the purposes of this experiment, was the finding of less forgetting following the interpolated tests over retention intervals of both 12 days [$F(1,30) = 29.85, MS_e = 7.75$] and 7 days [$F(1,28) = 24.46, MS_e = 7.17$].

In summary, the expectations and rehearsal of the subjects tended to lower the rate of forgetting (i.e., naive subjects tended to forget more over a given retention interval), as one would expect. This tendency was not consistent, however. For example, in Group 3T7, the naive subjects actually forgot less than the nonnaive subjects, over both the first 7 days and the last 7 days. The main point is that

the general trends of the data were the same for the complete groups as they were for the naive subjects.

Hypermnesia scores. As reported earlier, on Day 1 there was significant forgetting between Tests 1a and 1b as well as significant hypermnesia between Tests 1b and 1d, although there was no hypermnesia overall across the four tests on Day 1. These results support the view that some forgetting must occur before any hypermnesia is observed and that the absence of hypermnesia between Tests 1a and 1b in both experiments is due to the short interval (and consequent "recency effect") between the last study trial and Test 1a.

On each of the 3 days involving long-term recall, there was a significant increase in performance across the four tests [$F(3,84)$ = 4.36, MS_e = .42, $F(3,84)$ = 8.70, MS_e = .69, and $F(3,252)$ = 18.58, MS_e = .42, for retention intervals of 2 days, 7 days, and 14 days, respectively]. The extent of the hypermnesia over the tests can be seen in Figure 2 for the intermediate days and in Tables 1 and 2 for all conditions. Tests of individual means again showed that the amount of hypermnesia tended to decline over tests. For Days 3 and 8, only Tests 1 and 2 differed significantly. For Day 15, all pairs of tests differed among themselves, except for Tests 3 and 4. It should be pointed out that Group 3T2 showed no significant degree of hypermnesia on either Day 3 or Day 15.

These data were explored with additional analyses that used the amounts of hypermnesia rather than the recall performance as dependent variables. An analysis of the amounts of hypermnesia without the interpolated tests (Conditions 3T2-2, CT2-2, 3T7-7, CT7-7, 3-14, and C-14) showed that criterion learning produced more hypermnesia [$F(1,84)$ = 8.92, MS_e = 2.41], but neither retention interval nor the interaction had significant effects. An analysis of the data for the final tests (Conditions 3T2-14, CT2-14, 3T7-14, CT7-14, 3-14, and C-14) showed that only the level of learning × interpolated tests interaction had a significant effect [$F(2,84)$ = 4.03, MS_e = 1.41]. This interaction occurred because the interpolated tests had no overall effect for the three-trial condition, but attenuated hypermnesia for the criterial condition.

Next, the components of hypermnesia were analyzed. The analysis of the recovery data for Day 3 (Conditions 3T2 and CT2), Day 8 (Conditions 3T7 and CT7), and Day 15 (Conditions 3-14 and C-14), with level of learning and retention interval as between-subjects variables, showed that recovery declined over tests [$F(2,168)$ = 3.89, MS_e = 0.37] and was greater following criterion learning than following three study trials [$F(1,84)$ = 7.20, MS_e = 0.95]. There was, however, no significant effect of retention interval. The analysis of the forgetting data showed no significant effects, except that intertest forgetting declined across tests [$F(2,168)$ = 4.07, MS_e = 0.20]. The forgetting declined only between Tests 3 and 4. The effects of the intermediate tests on the final tests were analyzed by examining the data for Conditions 3T2-14, CT2-14, 3T7-14, CT7-14, 3-14, and C-14. This analysis

showed that the interpolated tests attenuated the amount of recovery [$F(1,88)$ = 5.25, MS_e = 0.68]. There was also a decline in recovery over tests [$F(2,176)$ = 3.56, MS_e = 0.40]. No factor had a significant effect on intertest forgetting for these groups.

Hypermnesia correlations. The coefficients of correlation between hypermnesia and Test 1 recall, cumulative recall, and the difference between these two measures were calculated and are shown in Table 3 for each condition. These data were analyzed statistically in the same manner as the corresponding data for Experiment 1 and showed a generally similar pattern. The criterion condition produced a significant overall correlational effect between hypermnesia and C_wR-T1, along with Test 1 recall [$\chi^2(10)$ = 33.76]; both learning conditions produced strong positive coefficients with C_sR-T1 [$\chi^2(10)$ > 66]. However, unlike Experiment 1, there were no significant correlations between hypermnesia and either measure of cumulative recall for the noncriterion learning conditions. In addition, the overall effect of the correlations for C_wR-T1 did not quite reach statistical significance for the three-trial conditions [$\chi^2(10)$ = 17.23], although they did if the anomalous 3T2-2 condition was omitted [$\chi^2(8)$ = 16.52]. The correlations between hypermnesia and the various measures collapsed across conditions were −.11, .14, .94, .14, and .36 for T1, C_sR, C_sR-T1, C_wR, and C_wR-T1, respectively. Only the third and fifth correlations were statistically significant (*p*s < .001).

Hypermnesia summary. The hypermnesia results largely replicated what was found in Experiment 1. Hypermnesia did not occur between Tests 1a and 1b on Day 1, but did occur when the retention interval was 2 or more days. Once again, retention interval and interpolated tests had no consistent effects on hypermnesia overall. However, recovery tended to decrease over tests and the interpolated tests tended to attenuate the amount of recovery found on the final tests. In addition, the best consistent, nontrivial predictor of the amount of hypermnesia was the difference between cumulative recall across the whole experiment and Test 1 recall. However, there were some differences from Experiment 1. Hypermnesia occurred on Day 1 between Tests 1b and 1d. Beyond Day 1, intertest forgetting tended to decline over tests, although this was confined to the last two tests only. Also, recovery did not decline significantly over retention interval. Finally, the correlations between hypermnesia and cumulative recall were not significant for the noncriterion conditions.

GENERAL DISCUSSION

The primary results of the two experiments are quite clear. The introduction of interpolated tests 2 or 7 days after study slowed down the rate of forgetting. This result generalizes from cued recall to free recall the conclusions of Runquist (1983, 1986a, 1986b) rather than those of Slamecka and Katsaiti (1988). In failing to replicate Slamecka and Katsaiti, the result here lends credence

to the view of Ballard (cited in Payne, 1987) that testing has a greater effect on the retrieved information when it occurs after a longer retention interval. With the exception of Group 1T in Experiment 1, the six groups that were given extra tests in the two experiments forgot far less over the relevant intervals than did their untested counterparts. The performance by Group 1T may also reflect attenuated forgetting but, as argued previously, their data cannot be unambiguously interpreted because the corresponding control condition (Group 1) showed no significant forgetting beyond Day 3. The view that the lack of forgetting in Group 1 was due to a floor effect was supported by the results of Group 3 in Experiment 2.

Questions that arise are: Why should interpolated tests attenuate rate of forgetting? Why should these extra tests do anything beyond raising the performance level (i.e., produce hypermnesia) and perhaps have a short-term effect on rate of forgetting, as found by Slamecka and Katsaiti (1988)? Explanations have been suggested in terms of enhancement of retrieval operations or the establishment of additional contexts (again, see Runquist, 1983, 1986a, 1986b). As a speculative example of the latter type of explanation, one might suggest that each successful recall of a word encodes that word in a new context and thereby increases the probability of recall of that word on a later test. Subjects who are not given extra tests must rely only on the traces of the earlier encodings.

This argument in turn raises the following question: Why should tests given 2 or 7 days after the study trials attenuate forgetting over the following 1 or 2 weeks, but tests given 30 sec after study have only a short-term effect, as found by Slamecka and Katsaiti (1988)? An answer is that the contexts for the interpolated tests given after 2 or more days are quite distinct from the study context, at least in temporal terms (see Glenberg, 1979). The same cannot be said for the contexts of tests given after 30 sec. While the latter may remain functionally distinct over an interval of a day or so, they will tend to blend in with the study context after intervals of several days. A corollary to this view is that the contexts of the interpolated tests here would eventually lose their distinctiveness and then the rate of forgetting under these conditions could, in theory, equal that of the nontested conditions. In practice, though, the retention intervals necessary to bring about this situation following the interpolated tests may be so long that the performance of the nontested control subjects is invariant along the floor.

Turning to additional findings, there were no differences in the rate of forgetting after an interval of 2 days between subjects who learned to criterion and those who did not. This finding was contrary to the prediction based on the results of Slamecka and Katsaiti (1988) and indicates that the overlearning induced here by criterion learning is not the same as the overlearning induced by their extra tests on Day 1. It also indicates that overlearning trials do not continue to add to the "strength" of a memory trace. Rather, it supports in a general way the view that acquisition is a stage process, such that once a memory reaches the final learned state, further study does not affect it (see, e.g., Brainerd, Howe, & Kingma, 1982).

Also, there were no differences in rate of forgetting between criterion and noncriterion conditions over the period from Day 3 to the final session 12 or 14 days later. This conclusion is based on the absence of a significant level of learning × retention interval interaction in Experiment 2 (and the assumption that the interaction in Experiment 1 resulted from the floor effects on Group 1). This finding extends the results of Slamecka and McElree (1983) to conditions of learning to a strict criterion.

Regarding the hypermnesia data, the absence of hypermnesia on Day 1 here, and in Rose and Howe (1989), was almost certainly due to the relatively short interval (1 min) between the last study trial and the first test trial (Test 1a). This suggestion is supported by the finding in Experiment 2 of a significant increase in recall from Test 1b to 1d following significant forgetting from Test 1a to 1b, and implies that some forgetting must occur before hypermnesia occurs.

It was predicted on the basis of the cumulative recall hypothesis that hypermnesia would be greater following criterion learning and less following longer retention intervals. In fact, level of learning and length of interval had inconsistent effects on the amount of hypermnesia. Also, insofar as they affected hypermnesia at all, the independent variables, including serial order of the tests, affected recovery and not intertest forgetting. Payne (1987) suggests that explanations of hypermnesia take into account those factors that affect intertest forgetting. On the contrary, the results from the two experiments reported here indicate that intertest forgetting, unlike recovery, is largely invariant across tests, and plays only a small role in hypermnesia, at least when long retention intervals and more than one set of tests are used.

The fundamental prediction of the cumulative recall hypothesis is that there is a positive correlation between amount of hypermnesia and cumulative recall. This result was not generally found here, although it must be emphasized that the present conditions, with long retention intervals, criterion learning, and interpolated testing were considerably different from those that led to the formulation of the hypothesis. In the present experiments, criterion learning produced a significant negative correlation between hypermnesia and Test 1 recall and, with one borderline exception, both learning conditions produced significant positive (and nontrivial) correlations between hypermnesia and the difference between whole-experiment cumulative recall and Test 1 recall.

Both of these results are intuitively reasonable. Hypermnesia depends on the difference between what is potentially retrievable and what is currently retrievable. If this difference is significant, then there is the potential for a significant amount of hypermnesia on one or more succeeding tests (assuming, of course, that intertest forgetting is relatively small and does not increase over tests). The problem is to determine what is potentially retrievable. In the types of studies reviewed by Payne (1987),

the only estimate of this value is the cumulative recall over a set of tests as employed by Roediger et al. (1982) and Payne (1986). In studies of the type reported here, the best estimate of potential retrievability is the cumulative recall taken as the total number of unique words recalled across all tests given over the entire retention interval. The greater the difference between net recall on the initial test and the overall cumulative recall, the greater should be the amount of hypermnesia over a set of tests. For criterion learning, the overall cumulative recall is fixed, of course, and so hypermnesia increases as Test 1 recall decreases.

The upshot of these arguments, based on the hypermnesia results found in these two experiments, is that the literal form of the cumulative recall hypothesis is a restricted predictor of the amount of hypermnesia to be found over a set of tests, and is limited to situations where a noncriterial study session is followed by one set of recall tests. A more generally applicable predictor is the difference between overall cumulative recall and initial test recall, although this predictor admittedly is trivial with noncriterial learning followed by one set of tests. The basic problem for predicting hypermnesia is to determine, in a nontrivial manner, what is potentially retrievable under a given set of learning and retrieval conditions.

REFERENCES

ALLEN, A. G., MAHLER, W. A., & ESTES, W. K. (1969). Effects of recall tests on long-term retention of paired associates. *Journal of Verbal Learning & Verbal Behavior*, **8**, 463-470.

BRAINERD, C. J., HOWE, M. L., & KINGMA, J. (1982). An identifiable model of two-stage learning. *Journal of Mathematical Psychology*, **26**, 263-293.

GLENBERG, A. M. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Memory & Cognition*, **7**, 95-112.

HARRISON, N. S. (1979). *Understanding behavioral research*. Belmont, CA: Wadsworth.

MCDANIEL, M. A., & MASSON, M. E. J. (1985). Altering memory representation through retrieval. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **11**, 371-385.

PAYNE, D. G. (1986). Hypermnesia for pictures and words: Testing the recall level hypothesis. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **12**, 16-29.

PAYNE, D. G. (1987). Hypermnesia and reminiscence in recall: A historical and empirical review. *Psychological Review*, **101**, 5-27.

ROEDIGER, H. L., PAYNE, D. G., GILLESPIE, G. L., & LEAN, D. S. (1982). Hypermnesia as determined by level of recall. *Journal of Verbal Learning & Verbal Behavior*, **21**, 635-655.

ROSE, R. J., & HOWE, M. L. (1989, June). *Effects of spacing, modality, and categorization on acquisition and retention*. Poster presented at the annual convention of the Canadian Psychological Association, Halifax, N.S.

RUNQUIST, W. N. (1983). Some effects of remembering on forgetting. *Memory & Cognition*, **11**, 641-650.

RUNQUIST, W. N. (1986a). Changes in the rate of forgetting produced by recall tests. *Canadian Journal of Psychology*, **40**, 282-289.

RUNQUIST, W. N. (1986b). The effect of testing on the forgetting of related and unrelated associates. *Canadian Journal of Psychology*, **40**, 65-76.

SLAMECKA, N. J., & KATSAITI, L. T. (1988). Normal forgetting of verbal lists as a function of prior testing. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 716-727.

SLAMECKA, N. J., & MCELREE, B. (1983). Normal forgetting of verbal lists as a function of their degree of learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **9**, 384-397.

UNDERWOOD, B. J. (1964). Degree of learning and measurement of forgetting. *Journal of Verbal Learning & Verbal Behavior*, **3**, 112-129.