Hypermnesia in free recall and cued recall

DAVID G. PAYNE, HELENE A. HEMBROOKE, and JEFFREY S. ANASTASI
State University of New York, Binghamton, New York

In three experiments, categorized lists and both free recall and cued recall tests were used to examine hypermnesia. In Experiment 1, materials were drawn from obvious and nonobvious categories in an attempt to vary the amount of relational processing at encoding. The study materials in Experiment 2 consisted of a long word list that comprised several exemplars from each of a number of common categories. In Experiment 3, a single exemplar was drawn from each of 45 categories. In each experiment, similar magnitudes of hypermnesia were obtained on free and cued recall tests. Examination of the specific items recalled across tests indicated that similar processes underlie the hypermnesic effect for both test conditions. Implications of the results for extant accounts of the hypermnesic effect are discussed. It is concluded that the dynamics of retrieval processes change in a systematic fashion across repeated tests and the retention interval following study and that an adequate account of the nature of these changes in retrieval dynamics is essential to our understanding of hypermnesia and related phenomena.

Most research on hypermnesia (improved memory performance associated with repeated testing or increased retention intervals) has employed either standard free recall tests or forced recall tests in which subjects are required to produce a specified number of items on each test. Generally speaking, hypermnesia is readily obtained with both test types (see, e.g., Erdelyi & Becker, 1974; Erdelyi & Kleinbard, 1978; Payne, 1986, 1987; Roediger & Payne, 1985; Roediger, Payne, Gillespie, & Lean, 1982). Efforts to provide a theoretical account of hypermnesia have generally been focused on the encoding conditions (e.g., imaginal encoding vs. nonimaginal encoding; Erdelyi & Becker, 1974; Erdelyi, Finkelstein, Herrell, Miller, & Thomas, 1976) and the parameters of the recall tests (e.g., number of repeated tests, length of tests; Roediger & Payne, 1982; Roediger & Thorpe, 1978) necessary to produce the effect. Recently, however, investigators have begun to expand the scope of inquiry by varying the type of tests used to assess memory performance (e.g., Otani & Hodge, 1991; Payne & Roediger, 1987) and the activities occurring between successive tests (e.g., Anastasi, Payne, Goodman, Lampkin, & Martinez, 1991; Shaw & Bekerian, 1991; Smith & Vela, 1991). In the present study, we examined the issue of whether hypermnesia is restricted to tests in which subjects are required to produce their own retrieval cues (e.g., free and forced recall tests) or whether the effect can also be obtained when subjects are provided with explicit retrieval cues (e.g., cued recall tests).

Understanding the role that retrieval processes play in producing hypermnesia is important because hypermnesia is readily obtained with free and forced recall measures but is not typically obtained with recognition tests. Payne and Roediger (1987) failed to obtain a hypermnesic effect with either yes/no or forced choice recognition tests under conditions in which a reliable hypermnesic effect was obtained with free recall tests. Otani and Hodge (1991) recently reported a similar failure to obtain a recognition hypermnesic effect. To our knowledge, only a single published experiment has yielded a recognition hypermnesic effect (Erdelyi & Stein, 1981). In Erdelyi and Stein's study, subjects were presented with a series of single-frame cartoons that consisted of a picture and a caption. For one half of these cartoons, the caption and figure were unrelated (i.e., the figures and captions had been re-paired across different cartoons). When subjects were given three successive yes/no recognition tests of their memory for either the picture or the caption, recognition memory performance (as indexed by d') improved only for the picture component of the unrelated pictures and captions. On the basis of this finding, along with a failure to obtain recognition hypermnesia with either pictures or captions from related cartoons or captions from unrelated cartoons, Erdelyi and Stein concluded that "for recognition hypermnesia to be obtained ... retrieval search must be a non-trivial component of the recognition task" (p. 30; see also Erdelyi, 1988, for a similar conclusion).

Payne and Roediger (1987) also concluded that retrieval demands play a critical role in producing hypermnesia. Payne and Roediger based their conclusion on the facts that (1) they failed to obtain recognition hypermnesia under conditions in which they found a significant free recall hypermnesic effect, and (2) Erdelyi and Stein's (1981) data showed that recognition hypermnesia was only obtained when the test imposed considerable retrieval de-

We thank M. H. Erdelyi, M. H. Hodge, Jr., A. W. Inhoff, H. Otani, A. F. Smith, M. J. Wenger, and especially J. I. Chumbley, for their comments on an earlier version of this manuscript. Portions of these data were reported at the 32nd Annual Meeting of the Psychonomic Society, San Francisco. Requests for reprints should be sent to D. G. Payne, Department of Psychology, SUNY-Binghamton, Binghamton, NY 13902-6000.

mands. Payne and Roediger extrapolated from these findings and suggested that hypermnnesia was likely to be obtained with tests that involved a large retrieval component (e.g., free recall tests) but not with recognition tests, since these tests typically bypass an effortful retrieval stage.¹ Furthermore, they also predicted that hypermnnesia was not likely to be easily obtained with cued recall tests "because typically powerful [retrieval] cues are given for studied items and those same cues would be repeated on later tests" (p. 160). If the retrieval cues provided on cued recall tests allow easy access to the target items (in a manner similar to that observed with recognition tests), cued recall tests would not seem to involve the type of retrieval processes necessary for producing a hypermnnesic effect.

Otani and Hodge (1991) have recently reported a cued recall hypermnnesic effect that conflicts with Payne and Roediger's (1987) prediction concerning cued recall tests. Otani and Hodge (Experiment 2) presented subjects with 36 word pairs, with the to-be-recalled item from each pair belonging to one of six different taxonomic categories. The cue word in each pair was a low-to-moderate associate of the target item. The subjects studied these items under one of three encoding conditions. In the relational processing condition, the subjects sorted the item pairs into categories based on the target items. Another group of subjects performed an individual item processing task, rating the pleasantness of each target word. The third group of subjects was given intentional learning instructions (i.e., they were simply told to try to memorize the word pairs).

After the target list was presented, the subjects were given three successive cued recall tests, with the associates of the targets as cues. Hypermnnesia was obtained only in the relational processing condition, with the other two conditions producing approximately equal recall levels across each test. In Experiment 3, Otani and Hodge (1991) used the same three encoding conditions, but this time the target items came from six loosely defined categories (e.g., big, round). Once again, hypermnnesia was obtained only in the relational processing condition.

On the basis of these results, Otani and Hodge (1991) suggested that "the occurrence of hypermnnesia in cued recall apparently requires the processing of relational information in the TBR [to-be-recalled] materials" (p. 114). They also speculated that the hypermnnesic effect obtained in cued recall following relational encoding may be attributable to the effective retrieval cues that are formed on the basis of the relational encoding.

The present research was designed to (1) examine the generality of the Otani and Hodge (1991) findings, (2) test their predictions regarding relational processing as a critical factor in cued recall hypermnnesia, and (3) resolve several empirical and theoretical issues concerning the Otani and Hodge study. One of these issues concerns the proper interpretation of Otani and Hodge's failure to obtain a hypermnnesic effect with their individual item processing and intentional encoding conditions. As Roediger et al. (1982; see also Payne, 1986) pointed out, the likelihood

of obtaining hypermnnesia depends critically on the difference between net recall on Test 1 (i.e., the number of items correctly recalled on Test 1) and the asymptotic level of cumulative recall, generally measured as the total number of unique items recalled across all tests. If net recall levels on Test 1 are near asymptote, there is little "room" for improvement, and hence it is unlikely that hypermnnesia will be obtained. (See Roediger and Challis, 1989, for a thorough discussion of this cumulative recall level account of hypermnnesia.)

It is important to note that asymptote has typically been operationally defined as the total number of unique items that subjects are able to recall under a given set of experimental conditions, and that this asymptote is typically less than the total number of items in the target set. For example, Roediger et al. (1982, Experiment 3) asked subjects to recall items from common categories on three tests. Subjects who were asked to recall U.S. presidents recalled, on the average, fewer than 20 of the (then 39) presidents. However, 88% of these recalled items were recovered during the first test period, and, as a result, this condition showed only a small hypermnnesic effect. Thus, in this condition, performance was near asymptote at the end of the first test, yet performance was nowhere close to the absolute ceiling (i.e., approximately 50% of all possible presidents were recalled). This illustrates the importance of discriminating between the empirical asymptote and the absolute possible asymptote (i.e., the total number of target items).

In the Otani and Hodge (1991) study, recall levels in the individual item and intentional conditions were quite close to asymptote at the end of Test 1. If we use cumulative recall across the three tests as the index of asymptotic cumulative recall (see Otani and Hodge, Table 1), then, by the end of Test 1, the subjects in the individual item and intentional conditions in Experiment 2 had recalled 89% and 91%, respectively, of all items recalled across all three tests. The corresponding values for Experiment 3 were 94% and 95%. In the relational processing condition—the condition in which hypermnnesia was observed—Test 1 net recall levels were the lowest of all three conditions in both Experiment 2 and Experiment 3 (75% and 90% of asymptote, respectively). Thus, one question that needs to be answered is whether, with the use of encoding conditions that do not require relational processing, hypermnnesia can be obtained with cued recall tests provided that performance is not approaching asymptote on Test 1.

A second issue concerning the Otani and Hodge (1991) study involves the retrieval cues used by subjects. As indicated earlier, Otani and Hodge drew the target items from only six categories. It is possible that the subjects recalled the category labels and used these labels as retrieval cues. If this was the case, it is unclear whether the functional retrieval cues used by subjects were (1) the experimenter-provided low-to-moderate associates, (2) the subject-generated category labels, or (3) a combination of (1) and (2). This point is important, because Payne and

Roediger's (1987) prediction of no hypermnnesia with cued recall tests assumes a strong relationship between retrieval cues and to-be-recalled items.

In order for Payne and Roediger's (1987) prediction to be tested adequately, it is important that the experimental procedures permit identification of the retrieval cues that subjects are utilizing. In Otani and Hodge's (1991) study, the subjects could have been using the experimenter-provided cues, in which case the cue-target relationship was not strong, contrary to Payne and Roediger's assumption in their prediction concerning cued recall hypermnnesia. Alternatively, if subjects were using category labels, then presumably these would provide strong cue-target relations and hence Otani and Hodge's data would provide evidence against Payne and Roediger's prediction. The present experiments were designed to provide tighter control over which retrieval cues subjects could use during testing and, more importantly, to use strong cue-target associates in some of the conditions.

Finally, Otani and Hodge (1991) did not include free recall conditions in their studies, and hence it is unclear whether a hypermnnesic effect would have been obtained with these materials and encoding conditions on repeated free recall tests. Since a free recall condition provides a useful baseline control condition against which a cued recall hypermnnesic effect can be compared, both cued and free recall tests were used in each of the present experiments.

The three experiments reported here provide evidence regarding each of these three issues. The general approach taken here was to present subjects with items from various types of categories and then administer either free recall tests or cued recall tests in which the category labels were provided as cues. In Experiment 1, we investigated whether or not cued recall hypermnnesia was limited to relational encoding conditions. In Experiments 2 and 3, we investigated the generality of hypermnnesia with free and cued recall tests. Of primary interest was whether or not, and under what conditions, hypermnnesia could be obtained with cued recall tests. Also of interest were the comparisons between the free and cued recall conditions, for these provide insight into the possible mechanisms affecting performance in repeated test paradigms. As we will see, considerable evidence indicates that changes in the dynamics of the retrieval processes that occur across tests play an important role in producing hypermnnesia. This retrieval dynamics view will be described fully after the experiments are presented.

EXPERIMENT 1

In Experiment 1, subjects studied a list of categorized items and then were given either three free recall tests or three cued recall tests with category labels as cues. For half the subjects, the categories were obvious categories and, on the basis of research by Hunt and Einstein (1981) and others (e.g., Klein, Loftus, Kihlstrom, & Aseron, 1989), we expected that these materials would induce relational pro-

cessing. The remaining subjects received items from nonobvious categories (e.g., green things), and, following Hunt and Einstein, we expected that these materials would induce item-specific processing. Finally, as each item was presented, subjects rated the item's pleasantness, an encoding condition that has been shown (e.g., by Hunt & Einstein) to promote individual item processing.

If Otani and Hodge's (1991) hypothesis that relational processing at encoding (see Otani and Hodge, p. 114) is important in producing cued recall hypermnnesia is correct, then hypermnnesia should only be found in the related list condition with cued recall tests. This prediction is based on the assumption that subjects are more likely to discern the categorical nature of the materials in the related list condition than in the unrelated list condition. On the other hand, in light of Otani and Hodge's findings of no cued recall hypermnnesia with item-specific or intentional encoding conditions, one might predict no cued recall hypermnnesia with *either* the related or the unrelated lists. This prediction is based on the fact that our encoding conditions were intended to foster individual item processing. We expected to obtain a free recall hypermnnesic effect with both sets of materials, since Payne (1987) has documented that free recall hypermnnesia with words is a fairly robust phenomenon. However, given the differences in materials between this experiment and the experiments reviewed by Payne, the free recall condition provided a test of whether any possible failure to obtain a cued recall hypermnnesic effect was due to the materials employed.

Finally, cumulative recall levels were examined to determine whether subjects approach asymptote by the end of Test 1. As indicated earlier, Otani and Hodge's (1991) failure to obtain a hypermnnesic effect in the individual item and intentional learning conditions might have been due to functional ceiling effects. If a ceiling effect was responsible for the null findings reported by Otani and Hodge, and if cued recall hypermnnesia is a reliable phenomenon, then provided that Test 1 recall levels were not close to ceiling, a cued recall hypermnnesic effect might be obtained in the unrelated list condition, which did not involve relational processing.

Method

Subjects and Design. The subjects were 80 undergraduate students enrolled in an introductory psychology course at the State University of New York at Binghamton who participated in partial fulfillment of a course requirement. The experiment employed a 2 (list type: related vs. unrelated) $\times$ 2 (test type: free recall vs. cued recall) $\times$ 3 (test: 1, 2, 3) mixed-factorial design. List type and test type were manipulated between subjects, and test was varied within subjects. The subjects were randomly assigned to one of the four experimental conditions and were tested in groups of 6 or fewer, with sessions lasting approximately 30 min.

Materials. Two sets of stimuli were derived from Hunt and Einstein's (1981) obvious and nonobvious category lists. Each list consisted of 48 items, with 8 items per category within each list. Each block of six items contained one item from each category, with order of items within each block determined randomly. Words on the related list were from six categories that were obvious and easily dis-

cernible (sports, fruits, clothing, animals, countries, and musical instruments). The unrelated list comprised words from six nonobvious categories (things that fly, green things, liquids, wooden things, women's things, and noisy things). For cued recall tests, a random ordering of the category labels was printed at the top of the sheet. The same ordering of category labels was used on each test.

Procedure. The subjects were informed that they would be presented with a list of words, and that their memory for those words would be tested. (The nature of the memory test was not specified.) In addition, the subjects were asked to rate the pleasantness of each word as it appeared, using a 7-point Likert-type scale, with 1 being *very unpleasant*, and 7 *very pleasant*. This rating procedure was included to ensure that subjects performed item-specific processing at encoding (cf. Hunt & Einstein, 1981).

The words were projected with a slide projector at a rate of 5 sec per word. After all the words were presented, the subjects' rating sheets were collected and the first of three recall tests was administered. The subjects in the free recall condition were instructed to recall, in any order, as many words as possible. The subjects in the cued recall conditions were provided with the response sheets with the names of each of the categories at the top. At 1-min intervals throughout each test, subjects were instructed to draw a line beneath the last word they had written. Each recall test lasted 6 min.

After this initial test, the subjects' test sheets were collected and new tests were distributed. The subjects were read a brief (about 1-min) set of instructions that explicitly asked them to attempt to improve their performance by recalling as many words from the prior test(s) as possible, while at the same time continuing to try to recall as many previously unrecalled words as possible. Test 2 began immediately after these instructions were given. Test 3 followed the same general procedures as did Test 2, with the exception that subjects were told that Test 3 was the last test.

Results and Discussion

Presented in Table 1 are the mean number of items recalled (i.e., net recall), on each test for each condition. All four conditions showed a considerable hypermnesic effect, with performance improving across each successive test in all conditions. Replicating previous studies employing these materials and similar encoding conditions (e.g., Klein et al., 1989), recall levels were higher in the related list condition than in the unrelated list condition. Finally, for both the related and unrelated conditions, the magnitude of the hypermnesic effect (i.e., the improvement in net recall from Test 1 to Test 3) was approximately equal in the free recall and cued recall test conditions.

These conclusions were supported by the results of a 2 (list type: related vs. unrelated) $\times$ 2 (test type: cued

recall vs. free recall) $\times$ 3 (Test 1, 2, or 3) mixed factor analysis of variance (ANOVA). There were significant main effects of test [$F(1, 152) = 74.6$, $MS_e = 4.3$] and list type [$F(1, 76) = 84.3$, $MS_e = 57.5$]. (An alpha level of .05 was used in all statistical tests.) The slight difference in performance levels for the cued and free recall conditions did not approach significance ($F < 1.0$). Furthermore, there was not a significant difference in the magnitude of the hypermnesic effect for the free recall and cued recall conditions [interaction $F(2, 152) = 1.88$, $MS_e = 4.3$, $p > .15$].

This same pattern of results concerning the effects of tests and test type was obtained in separate analyses of the data from the related and unrelated conditions as well as separate analyses on the data from the cued recall and free recall conditions. The net recall data thus do not appear to support either Payne and Roediger's (1987) prediction of no hypermnesia with cued recall or Otani and Hodge's (1991) hypothesis that hypermnesia will be obtained in cued recall tests only following relational encoding. Hypermnesia was found in both of the cued recall conditions, and one of these conditions employed materials (an unrelated list) and encoding conditions (pleasantness ratings) that were unlikely to promote relational processing.

It was suggested earlier that Otani and Hodge (1991) failed to obtain a cued recall hypermnesia effect in their individual item and intentional conditions because performance levels may have been approaching asymptote on Test 1. If this interpretation of Otani and Hodge's results is tenable, one would expect that in the present experiment Test 1 net recall levels should be far from asymptote. To test this, cumulative recall levels across the three tests were examined.

Cumulative recall across the three tests was calculated by giving subjects credit for recalling an item only the first time it was recalled and ignoring any subsequent recall(s) or failure(s) to recall the item. Of primary interest here was the proportion of all items that would be recovered across the entire time of testing that were recalled on Test 1. The mean cumulative recall levels across the three tests for each of the four conditions are presented in Figure 1. There are two important points to note with regard to these data. First, it is clear that performance levels at the end of Test 1 were not approaching asymptote. For the related list condition, the cumulative recall levels at the end of Test 1 were 74.6% and 73.8% of the cumulative recall level at the end of Test 3 for the cued recall and free recall conditions, respectively. The corresponding data for the unrelated list were 83.4% and 75.5%. These data, along with the net recall data, indicate that when recall levels are not approaching asymptote at the end of Test 1, hypermnesia can be obtained with both free and cued recall tests. The current results were obtained when recall levels were not approaching asymptote on Test 1, indicating that Otani and Hodge's (1991) failure to find hypermnesia may have been the result of ceiling effects in their item-specific conditions. In

Table 1
Experiment 1: Mean Net Recall on Tests 1-3 and
Improvement Across Tests for the Free Recall and Cued
Recall Tests in the Related and Unrelated List Conditions

Test Type	Test			Test 3 – Test 1
	1	2	3	
Related List				
Free recall	26.20	28.15	31.30	5.10
Cued recall	28.15	29.15	31.30	3.15
Unrelated List				
Free recall	17.50	20.15	21.70	4.20
Cued recall	18.35	20.65	21.95	3.60

Figure 1. Mean cumulative recall level for the free and cued recall test conditions in Experiment 1.

the General Discussion, we will consider the impact of these data for Payne and Roediger's (1987) predictions.

A second point to note regarding the cumulative recall data is the differential pattern of cumulative recall levels for the free recall and cued recall tests in the related and unrelated test conditions. In the unrelated test condition, the cued recall condition showed a slight advantage across all three tests. In contrast, in the related list condition, cued recall exceeded free recall initially, but by the end of Test 3, cumulative recall in the free recall condition exceeded that observed in the cued recall condition.

Cumulative recall was analyzed by comparing the total number of items recovered at the end of each test. As indicated in the lower panel of Figure 1, the rate of recovery was equivalent for the two test conditions for the unrelated list [interaction $F(2,76) = .8$, $MS_e = 2.8$]. In the related list condition (see the upper panel), there was a

significant test type $\times$ test interaction [$F(2,76) = 5.5$, $MS_e = 3.9$]. None of the other effects approached significance. Additional support for the differential patterns of results was obtained by analyzing cumulative recall for each minute of testing. A 2 (list type) $\times$ 2 (test type) $\times$ 18 (minutes) ANOVA with the cumulative number of items recalled in each minute as the dependent variable was conducted. There was a significant main effect of list type [$F(1,76) = 102.5$, $MS_e = 293.4$], reflecting the fact that subjects in the related list condition recovered more items than did subjects in the unrelated list condition. The main effect of minute was, of course, significant [$F(17,1,292) = 512.3$, $MS_e = 3.09$]. More importantly, the two-way interactions of list type $\times$ minute [$F(17,1,292) = 9.9$, $MS_e = 3.09$] and test type $\times$ minute [$F(17,1,292) = 2.6$, $MS_e = 3.09$], as well as the three-way (list type $\times$ test type $\times$ minutes) interaction [$F(17,1,292) = 4.8$, $MS_e = 3.09$] were all significant.

To further examine the three-way interaction, separate 2 (test type) $\times$ 18 (minutes) ANOVAs for the related and unrelated conditions were conducted. In the unrelated list condition, there was no evidence of a test type $\times$ minute interaction [$F(17,646) = 1.2$, $MS_e = 2.5$], whereas the interaction was significant in the related list condition [$F(17,646) = 5.5$, $MS_e = 3.7$].

We propose that the difference in the cumulative recall levels for the free and cued recall groups in the related list condition can best be understood by considering the retrieval cues available in these two conditions. We assume that in addition to general contextual cues, there are two other types of cues that subjects can use to guide retrieval: category labels and the list items that are successfully recalled. These assumptions are consistent with extant models of recall performance (e.g., Mensink & Raaijmakers, 1988; Raaijmakers & Shiffrin, 1980).

It is quite likely that subjects in the related list condition identified the categorical nature of the items. In the unrelated list condition, it is unlikely that subjects spontaneously identified the categories represented in the list during the list presentation. Thus, the encoding specificity hypothesis (Tulving & Thomson, 1973) would predict that the unrelated list category labels would not be very effective retrieval cues since these cues are not present at the time of encoding. On the other hand, the related list cues were implicitly present and should be effective retrieval cues. Besides encoding specificity effects, the strong preexperimental association between the category labels and exemplars in the related list condition, but not in the unrelated list condition, would increase the likelihood of exemplar generation during recall.

In the cued recall test conditions, on each test there was a set of cues provided by the test that were constant across tests (i.e., the category labels), whereas in the free recall test, the retrieval cues that were available (list items and/or category labels) had to be generated by the subject. Thus, for the related list condition in which strongly related category labels were present during cued recall, there was high similarity in the retrieval cues across tests. In con-

trast, the specific items recalled across free recall tests should vary more than the items recalled across cued recall tests because, among other things, subjects tend to recall the items in slightly different orders on each test. This would change the nature of the functional retrieval probes across tests. If previously recalled items served as retrieval probes (cf. Raaijmakers & Shiffrin, 1980), then, in the free recall condition, these items exerted a powerful influence over which items were recalled next. In contrast, category labels likely played a major role in determining which items were recalled on cued recall tests. Since the cumulative recall measure only takes into account the first time an item is retrieved and disregards any subsequent retrieval failure(s) or repeated recall(s) of the item, the related-list free recall condition would be expected to have a higher level of cumulative recall, since on each test the subject generated somewhat different sets of retrieval cues that likely allowed access to different sets of items. This increase in item gains in the free recall condition was probably offset by an increase in item losses between tests, assuming that the retrieval cues that allow subjects to gain access to new items were different from the retrieval cues that allowed access to items recalled on previous tests.

In the unrelated list condition, the category cues were not strongly related and thus the retrieval processes in free and cued recall should be very similar. Finally, note that the net recall measures for both the related and the unrelated list conditions produced no evidence of a test type $\times$ tests interaction. According to the hypotheses outlined above, these null results are attributable to two different patterns of item recall across tests. In the related list condition, free recall tests should lead to increases in both item gains and item losses (relative to the cued recall condition), whereas in the unrelated list condition, the net recall data presumably reflect equivalent rates of item gains and losses in the free and cued recall conditions.

Evidence in support of the preceding hypotheses was sought in a set of analyses of rates of intertest forgetting (items recalled on test n but not on test $n+1$) and item recovery (items recalled on test $n+1$ but not on test n). Presented in Table 2 are the mean number of items forgotten and recovered between successive tests for each of the four groups. Note that, because of the differential net recall levels in the related and unrelated list conditions, the opportunities for item recoveries/losses in these two conditions vary greatly. As a result, the recovery/loss data for these conditions were analyzed separately in 2 (test type: free recall vs. cued recall) $\times$ 2 (test pairs: Test 1-Test 2 vs. Test 2-Test 3) $\times$ 2 (component of recall: forgetting vs. recovery) ANOVAs.

In both the related and the unrelated list conditions, the rates of item recovery exceeded intertest forgetting [$F(1,38) = 39.0$, $MS_e = 4.2$, and $F(1,38) = 86.3$, $MS_e = 1.8$, respectively]. More importantly, however, was the differential effect of test type in the two list conditions. In the unrelated list condition, the item fluctuation rate (i.e., losses and recoveries) was equivalent in

Table 2
Experiment 1: Mean Number of Items
Forgotten and Recovered Between Successive Tests

Test Type	Item Type	Test Trials		
		1-2	2-3	<i>M</i>
	Related List			
Free recall	Forgotten	2.90	2.25	2.58
	Recovered	4.80	5.30	5.05
Cued recall	Forgotten	2.00	1.30	1.65
	Recovered	3.00	3.45	3.22
	Unrelated List			
Free recall	Forgotten	1.20	1.25	1.22
	Recovered	3.85	2.80	3.32
Cued recall	Forgotten	1.29	1.85	1.57
	Recovered	3.50	3.15	3.32

the free recall and cued recall conditions [$F(1,38) < 1.0$]. For the related list condition, the item fluctuation rate was much higher in the free recall condition than in the cued recall condition [$F(1,38) = 21.5$, $MS_e = 4.2$], exactly as predicted by the foregoing analysis.

This pattern of results is in line with the hypotheses discussed above, and it suggests that the specific retrieval processes involved in the various testing conditions play a major role in determining not only the presence or absence of the hypermnesic effect but also influence the specific pattern of items recalled across successive tests. This conclusion will be considered further in the General Discussion.

EXPERIMENT 2

The results of Experiment 1 provide evidence relevant to two theoretical positions concerning cued recall hypermnesia. First, these data are contrary to Payne and Roediger's (1987) prediction that hypermnesia would not be obtained with cued recall tests. Second, given that the unrelated list and pleasantness encoding manipulations were designed to minimize relational processing, the results from Experiment 1 appear to be contrary to Otani and Hodge's (1991) suggestion that cued recall hypermnesia is limited to conditions involving relational processing. However, taken together, the results of Experiment 1 and the Otani and Hodge study suggest that cued recall hypermnesia may be a robust phenomenon that occurs under a range of encoding conditions, provided that performance levels do not approach asymptote during the initial test period. Experiment 2 provided additional evidence regarding the conditions under which a cued recall hypermnesic effect can be obtained.

A second issue addressed in Experiment 2 was whether the processes responsible for producing hypermnesia are similar in free recall and cued recall tests. In Experiment 2, we obtained indirect evidence regarding this issue by examining the manner in which improvements in net recall are produced across tests. When categorized lists are used, subjects can improve their net recall across

tests by recalling additional items from categories accessed on earlier tests and/or recalling items from categories not accessed on earlier tests. In Experiment 1, there were only a small number (six) of categories in each list and subjects typically recalled items from all categories on each test. Across conditions, the average number of categories represented in subjects' recall protocols ranged from 5.8 to 6.0 out of 6 possible categories. In order to eliminate this ceiling effect, in Experiment 2 we used a 100-item list that contained many more categories (20) than were used in Experiment 1.

The materials and study conditions of Experiment 2 were the same as those used by Payne (1986, Experiment 4). As in Experiment 1, half the subjects received three free recall tests and half received three cued recall tests with category label cues. Given Payne's (1986) findings, a hypermnesic effect in the free recall condition was expected. If the cued recall hypermnesic effect reflects processes similar to those responsible for the free recall hypermnesic effect, one would expect a parallel pattern of results (in terms of the number of items per category recalled and/or the number of items accessed across tests) in the free and cued recall test conditions.

Method

Subjects and Design. A 2 (list: A vs. B) $\times$ 2 (test type: free recall vs. cued recall) $\times$ 3 (test: 1, 2, or 3) mixed factorial design was used with list and test type manipulated between subjects, and test varied within subjects. Eighty-four undergraduate students were selected from the same source as in Experiment 1 and were assigned randomly to one of four groups ($n = 21$). The subjects were tested in groups of 6 or fewer in a single session lasting approximately 40 min.

Materials. The two word lists employed were those used in a previous study by Payne (1986, Experiment 4). Each list consisted of 100 words, 5 from each of 20 categories represented in the Battig and Montague (1969) category norms. The category instances were drawn from Frequency Positions 7–25 in the category norms, and the items' response frequencies were equated across List A ($M = 65.52$) and List B ($M = 65.53$). Two study lists were used to examine guessing, but, replicating Payne (1986), the guessing rates were extremely low and were not systematically related to either list or test type.

Category labels and instances were typed and mounted on slides. Category labels were typed in uppercase letters, while exemplars were in lowercase type. Exemplars of each category were presented in blocks of five following the presentation of their respective category label. The presentation order of category labels and their respective exemplars was determined randomly and was the same for both lists.

Procedure. The subjects were instructed that they would be presented with a list of words and that their recall for these words would be subsequently tested. The format of the list was explained, and the subjects were told that although they would be tested for their memory of the category instances only, they should pay attention to the category labels, because this would help them learn the items. A slide projector was used to present the words at a rate of 5 sec per word.

After the presentation of the list, the subjects were given an 8-min recall test. The subjects in the free recall conditions were simply instructed to recall as many words as possible, in any order. For the subjects in the cued recall conditions, category labels were listed in a random order at the top of the response sheets. (The same

ordering was used on each test.) These subjects were instructed to recall as many items as possible, regardless of order, and to use the category labels to help them remember individual items. At 1-min intervals throughout the tests, the subjects were prompted to draw a line underneath the last item recalled.

The initial recall test was followed by two additional 8-min recall tests. The procedures for these subsequent tests were similar to those used in Experiment 1. At the end of each test, the subjects' test sheets were collected and new tests were distributed. The instructions for Tests 2 and 3 explicitly asked subjects to attempt to improve their performance relative to the preceding test(s). The instructions for Test 3 indicated that it would be the last test.

Results and Discussion

Three dependent variables were used to evaluate the subjects' recall performance on each test: the number of items recalled on each test, the number of categories recalled per test, and the number of items per category (IPC) recalled on each test (see Table 3). Category recall was defined as recalling at least one item from the target category, and IPC measured how many items were recalled from accessed categories. These data were analyzed with three separate 2 (test type: free recall vs. cued recall) $\times$ 2 (list: A vs. B) $\times$ 3 (test: 1, 2, or 3) ANOVAs, one for each dependent measure.

Analysis of the net recall data indicated that a significant hypermnesic effect was obtained [$F(2, 160) = 95.5$, $MS_e = 13.7$]. There was also a significant effect of test type [$F(1, 80) = 8.8$, $MS_e = 614.4$]. The magnitude of the improvement in net recall scores was equivalent for free and cued recall [interaction $F(2, 160) = 1.2$, $MS_e = 13.7$], with simple effects tests revealing a significant main effect of test for both free recall [$F(2, 160) = 51.9$, $MS_e = 13.7$] and cued recall [$F(2, 160) = 44.9$, $MS_e = 13.7$]. Thus, in both Experiment 1 and Experiment 2, there was a significant improvement in net recall levels across the free and cued recall repeated tests, and there was no difference in the magnitude of this improvement.

The category recall and IPC measures indicate the manner in which the increase in net recall levels was achieved. Consider first the category recall measure. Cued recall

Table 3
Experiment 2: Mean Number of Words,
Categories, and Items Per Category Recalled Across
the Three Tests in the Free Recall and Cued Recall Conditions

Dependent Variable	Test			Test 3 – Test 1
	1	2	3	
Free Recall Test				
Words	31.36	35.52	39.57	8.21
Categories	11.50	12.48	13.81	2.31
Items/category	2.65	2.77	2.80	0.15
Overall group data*	2.73	2.85	2.86	
Cued Recall Test				
Words	40.55	45.74	48.00	7.45
Categories	15.79	17.02	17.29	1.50
Items/category	2.52	2.62	2.71	0.19
Overall group data*	2.57	2.69	2.78	

*Items/category data based on overall group data—that is, group mean net recall divided by group mean number of categories.

tests allowed subjects to access more categories than did the free recall tests [$F(1,80) = 50.8$, $MS_e = 2.5$]. Both groups showed an increase in the number of categories accessed across each of the three tests, although in the cued recall condition the improvement from Test 2 to Test 3 was somewhat smaller, perhaps reflecting a functional ceiling effect. The overall ANOVA of these data revealed a significant main effect of test [$F(2,160) = 55.6$, $MS_e = 1.4$] and a significant test $\times$ test type interaction [$F(2,160) = 4.7$, $MS_e = 1.4$]. The conclusion that the test $\times$ test type interaction reflects a functional ceiling effect in the cued recall condition is supported by the fact that there was no evidence of a test $\times$ test type interaction in an ANOVA of only the data from the first two tests ($F < 1.0$).

Payne (1986, Experiment 4) found that there was a slight (but not significant, $p = .11$) increase in the IPC measure across tests. A similar small but significant [$F(2,160) = 25.1$, $MS_e = .02$] increase was obtained in the present experiment. More importantly, the free and cued recall conditions yielded equivalent overall IPC levels ($F < 1.0$), and the magnitude of the increase across tests was the same for the two conditions ($F < 1.0$).

These results replicate and extend the findings from Experiment 1. Hypermnesia was obtained with both free and cued recall tests, using intentional learning instructions. The cued recall data stand in contrast to Otani and Hodge's (1991) failure to obtain a hypermnesic effect with intentional learning and items from common taxonomic categories. The results of Experiment 2 also extend the generality of the cued recall hypermnesic effect obtained in Experiment 1.

Another important point with regard to the category recall and IPC measures is that Experiment 2 demonstrated that the processes underlying the improvement in free and cued recall tests seem to be similar. Note that although this conclusion is based on a failure to reject the null hypothesis, it is not the case that there was no change in performance across tests. Both test conditions showed increases in net recall levels, the number of items per category, and the number of categories recalled. Also, replicating Payne (1986, Experiment 4), the increase in net recall levels was largely due to an increase in the number of categories accessed. In the present experiment, of the categories that were not accessed by subjects in Test 1, 27% of these were accessed by the end of Test 3 in the free recall condition and 36% in the cued recall condition. In contrast, the percentage increase in nonrecalled items per category accessed was only 6% and 8% in the free and cued recall conditions, respectively. Overall, then, these data suggest rather strongly that the hypermnesic effect is similar in both cued and free recall tests.

We also examined the rates of intertest forgetting and item recovery across tests in the free and cued recall conditions (see Table 4). These data were analyzed using a 2 (test type: free recall vs. cued recall) $\times$ 2 (component of recall: forgetting vs. recovery) $\times$ 2 (list: A vs. B) $\times$ 3 (test pairs: 1-2, 2-3) ANOVA. As expected, item recovery

Table 4
Experiment 2: Mean Number of Items
Forgotten and Recovered Between Successive Tests

Test Type	Item Type	Test Trials		
		1-2	2-3	<i>M</i>
Free recall	Forgotten	3.43	2.12	2.78
	Recovered	7.55	6.17	6.86
Cued recall	Forgotten	2.00	2.07	2.04
	Recovered	7.17	4.29	5.73

exceeded forgetting [$F(1,80) = 131.5$, $MS_e = 9.6$]. Item fluctuations were greater between Tests 1 and 2 than between Tests 2 and 3 [$F(1,80) = 20.1$, $MS_e = 7.9$]. There was also somewhat greater fluctuation in the specific items recalled across tests in the free recall condition than in the cued recall condition, although this difference did not reach significance ($p = .07$).

The interpretation of these main effects is complicated somewhat by the presence of two higher order interactions. First, there was a significant component of recall $\times$ test pair interaction [$F(1,80) = 4.0$, $MS_e = 12.0$], indicating that, averaged across test type, the decrease in the recovery rate across tests (7.36 to 5.23) was greater than the decrease in the intertest forgetting rate (2.72 to 2.10). However, this interaction must be viewed in light of the fact that there was a marginal three-way interaction of test type $\times$ test pair $\times$ component of recall [$F(1,80) = 3.63$, $MS_e = 12.0$, $p = .06$]. This interaction indicates that the change in the rate of recovery and forgetting across tests was different in the free recall and cued recall conditions.

To further explore this three-way interaction, we examined the data from the cued recall and free recall conditions in separate 2 (component of recall: forgetting vs. recovery) $\times$ 2 (list: A vs. B) $\times$ 3 (test pairs: 1-2, 2-3) ANOVAs. These analyses revealed the same main effects as in the overall ANOVA (i.e., component of recall, test pair). The only difference in the two ANOVAs was that there was a significant test pair $\times$ component interaction in the cued recall condition [$F(1,40) = 8.24$, $MS_e = 11.11$] but not in the free recall condition [$F(1,40) = .01$, $MS_e = 12.91$]. In the cued recall condition, the rate of intertest forgetting was constant across the test pairs, whereas the recovery rate was higher between Tests 1 and 2 than between Tests 2 and 3. In the free recall condition, the decrease in forgetting rates across test pairs (1.31) was equivalent to the change in the recovery rate (1.38). Put another way, in the cued recall condition there was a large decrease in the recovery rate across test pairs (2.88 items), whereas in the free recall condition the recovery rate decreased much less (1.38 items) between successive test pairs.

This differential pattern in the rates of item recoveries across the two test conditions is understandable in light of the differences in the retrieval cues available in the two conditions—namely, strong retrieval cues in the cued recall condition and no externally presented cues in the free

recall condition. This issue of differences in retrieval cues will be pursued further in the General Discussion.

EXPERIMENT 3

The results of Experiments 1 and 2 contradict Payne and Roediger's (1987) prediction of no hypermnnesia when strong cues are used as the retrieval cues. Experiment 3 provided another test of this prediction but corrected for what may have been a design flaw in the first two experiments. This experiment represents a conceptual replication of Experiment 2, in that we once again employed an intentional learning manipulation and category instances as the to-be-recalled items.

In Experiments 1 and 2, there were several (five and eight, respectively) instances from each category represented in the list. It is possible that cue overload (Watkins & Watkins, 1975) decreased the overall effectiveness of the category labels as retrieval cues, since a number of target items were associated with each category cue. It is possible that cue overload reduced the effective strength of the category cues, thereby leading to only an apparent inconsistency with Payne and Roediger's (1987) hypothesis. To test this possibility, we presented subjects with a large number of categories and only one instance from each category. If the cue overload account is accurate, then a hypermnestic effect should not be found with the cued recall tests. To further ensure that each cue provided on the recall tests was a strong associate of the target item, subjects rated the category typicality of each item as it was presented. Rating typicality should force subjects to attend to the items' category-specific features and thus make the category label a strong cue at test.

Method

Subjects and Design. A total of 28 undergraduate students participated in the experiment. Nineteen students participated as part of a classroom demonstration, and 9 students were drawn from the same source as in Experiment 1. Approximately half the students from each subject pool were assigned to the cued recall condition and half to the free recall condition. The experiment was thus a 2 (test type: free recall vs. cued recall) $\times$ 3 (test: 1, 2, or 3) mixed factorial design.

Materials. The to-be-recalled list consisted of 45 words, 1 from each of 45 different categories in the Battig and Montague (1969) norms. All category instances were drawn from Frequency Positions 7-18 ($M = 11.3$) in the category norms. Category names and instances were typed on an overhead transparency, with category labels in uppercase letters on the left and category instances in lowercase letters on the right. Each of the three cued recall tests consisted of a different random ordering of the category labels arranged in two columns on the test sheets. All subjects in the cued recall condition received a different cued recall form on each test.

Procedure. The subjects were instructed that they would be studying a long list of pairs of items and that each pair would contain a category name in uppercase letters and an instance from the category in lowercase letters. The subjects were informed that their memory would be tested for the category instances only, but that they should study the category names as well since these might help them to remember the category instances. The type of memory test was not specified. The subjects were also instructed that as each item pair was presented, they were to rate how typical/representa-

tive each category instance was of the category on a scale from 1 (*very atypical*) to 7 (*very typical*).

After these instructions, the subjects went through a brief practice list to familiarize themselves with the study procedure. For both the practice and the critical lists, the study items were presented via overhead transparency with a sliding mask used to control item presentation. After the practice, the subjects' questions were answered and then the critical list was presented at an 8-sec rate. After the list was presented, the subjects recalled the states of the United States for 1 min as a distracter task. Following this distracter task, the first recall test was distributed. The subjects were instructed that some of their tests contained the category labels (cued recall condition) and that these names should be used to help them recall the category instances. The subjects were instructed to write the category instances alongside the appropriate category labels if these were given, but if their recall test did not contain the category labels (free recall condition), they should simply recall as many category instances as possible and write these down on the test sheets in any order. Test 1 began immediately after these instructions and lasted for 4 min. The procedures for Tests 2 and 3 were similar to those in the previous experiments, including (1) the request that subjects attempt to improve their recall performance on each test, and (2) the notification that Test 3 was the last test.

Results and Discussion

Table 5 shows the mean number of items recalled on each test for the two conditions. These data were analyzed with a 2 (test condition: free recall vs. cued recall) $\times$ 3 (test: 1, 2, or 3) mixed factor ANOVA. As expected, subjects recalled more items in the cued recall condition than in the free recall condition [$F(1,26) = 320.1$, $MS_e = 46.4$]. More importantly, there was also a significant main effect for test, indicating a hypermnestic effect [$F(2,52) = 26.7$, $MS_e = 2.5$]. Although there was a larger improvement in the free recall condition than in the cued recall condition [interaction $F(2,52) = 3.73$], this effect is likely due to a ceiling effect in the cued recall condition. Simple effects tests indicated that for both the free and cued recall test conditions there was a significant improvement in net recall levels across the three tests [$F(2,52) = 25.07$, $MS_e = 2.5$, and $F(2,52) = 5.31$, $MS_e = 2.5$, respectively].

Presented in Table 6 are the mean rates of intertest forgetting and item recovery across tests. These data were analyzed with a 2 (test type: free recall vs. cued recall) $\times$ 2 (component of recall: forgetting vs. recovery) $\times$ 3 (test pairs: 1-2, 2-3) ANOVA.² As in Experiment 2, the item gain rate exceeded the forgetting rate and there was greater item fluctuation in the free recall condition than the cued recall condition (both $ps < .001$). There was also an overall greater rate of forgetting and recovery between Tests 1 and 2 than between Tests 2 and 3. The com-

Table 5
Experiment 3: Mean Number of Items
Recalled on Tests 1-3 and Improvement Across
Tests for the Free Recall and Cued Recall Test Conditions

Test Type	Test			Test 3 - Test 1
	1	2	3	
Free recall	12.21	15.36	16.28	4.07
Cued recall	40.14	41.43	42.07	1.93

Table 6
Experiment 3: Mean Number of Items
Forgotten and Recovered Between Successive Tests

Test Type	Item Type	Test Trials		<i>M</i>
		1-2	2-3	
Free recall	Forgotten	1.28	1.71	1.50
	Recovered	4.43	2.64	3.54
Cued recall	Forgotten	0.00	0.00	0.00
	Recovered	1.28	0.64	0.96

ponent of recall $\times$ test pair interaction indicates that, overall, there was greater item fluctuation between Tests 1 and 2 than Tests 2 and 3. The significant component of recall $\times$ test type interaction is due to the fact that in the cued recall condition there was essentially no forgetting between tests. This is consistent with the view that the category labels served as strong retrieval cues.

GENERAL DISCUSSION

A primary goal of this study was to determine under what conditions a cued recall hypermnesic effect could be obtained. The conditions necessary for producing a cued recall hypermnesic effect are important for our understanding of hypermnnesia and related phenomena (e.g., item gains, item losses across tests) as well as the more general issue of how attempts to retrieve information affect the memory system.

Otani and Hodge (1991) reported the first case of a cued recall hypermnesic effect and their results suggested that the effect may be limited to certain conditions involving relational processing at encoding. The present experiments extend this important finding by demonstrating that cued recall hypermnnesia can be obtained under a variety of encoding conditions and across a range of materials.

One point on which the results of the present study diverge from those of Otani and Hodge (1991) regards whether relational processing at encoding is critical in producing a cued recall hypermnesic effect. In the Otani and Hodge study, cued recall hypermnnesia was restricted to conditions involving relational processing. However, in Experiment 1, a cued recall hypermnesic effect was obtained under conditions that were not likely to involve relational processing at encoding since unrelated items and an encoding task (pleasantness rating) that induced primarily individual item processing (Hunt & Einstein, 1981) were used. Furthermore, this cued recall hypermnesic effect was similar in magnitude to that observed with free recall tests. These cued recall results, along with the cumulative recall data reported here and in the Otani and Hodge study, suggest that the differential pattern of hypermnesic findings across their encoding conditions may have reflected differences in the degree to which the conditions were approaching asymptote during the first test. Final determination of the role of relational processing in cued recall hypermnnesia will await additional experiments in which performance levels are systematically

varied in conditions involving relational versus individual item processing.

The cued recall hypermnesic effect reported here is also inconsistent with Payne and Roediger's (1987) prediction that hypermnnesia should not be obtained with cued recall tests. It is possible, however, to argue that the retrieval cues used in the present experiments were not related strongly enough to the target items and hence did not provide an adequate test of Payne and Roediger's prediction. For example, in Experiment 3, only a single exemplar from each category was used, but the exemplars were not the most dominant in their respective categories. Perhaps if more dominant exemplars had been used, a cued recall hypermnesic effect would not have been obtained. Note, however, that the recall levels in the cued recall condition of Experiment 3 were very high; subjects recalled 90% of the target items in Test 1. If more dominant exemplars had been used, it is quite likely that the subjects' performance would have been even higher, and a failure to obtain a cued recall hypermnesic effect would thus have been uninterpretable, because of ceiling effects.

It is possible, of course, that with different materials, study conditions, retrieval cues, and/or retention intervals it may be possible to obtain a free recall hypermnesic effect but not a cued recall hypermnesic effect. Indeed, such a demonstration would represent an important boundary condition for the hypermnnesia phenomenon. The fact remains, however, that in studies done with cued recall tests, when ceiling effects do not prevent an improvement in recall, hypermnnesia is the rule rather than the exception. These studies thus demonstrate that cued recall hypermnnesia is a reliable phenomenon, and they extend the range of conditions under which the effect is observed.

A Retrieval Dynamics Account of Hypermnnesia

A major challenge remaining for hypermnnesia researchers is to provide an adequate theoretical explication of hypermnnesia. While early hypermnnesia studies tended to concentrate on the changes in net recall levels observed across successive tests, more recent studies have focused on the patterns of item gains and losses across tests (e.g., Klein et al., 1989; Payne, 1986; Smith & Vela, 1991) and the temporal characteristics of recall across the test period(s) (e.g., Roediger & Payne, 1982; Roediger et al., 1982; Roediger & Thorpe, 1978). These two sets of factors—fluctuations in item recall across tests, and the temporal characteristics of retrieval—may point to the possibility of a more detailed explanation of hypermnnesia.

We will focus our attention here on the role that retrieval processes play in producing hypermnnesia, because these have been implicated in the major theoretical accounts of the phenomenon. We will first briefly review these three accounts and then demonstrate how the general concepts embodied in these accounts can be used to interpret the results from the present experiments as well as some other recently reported data. To foreshadow a bit, we will see that each of these accounts stresses the importance of changes in the dynamics of retrieval over time.

Erdelyi and Becker (1974) presented the first modern account of hypermnnesia. They obtained hypermnnesia for pictures but not for words, and hence their account focused on hypermnnesia for pictorial items. In subsequent studies, the reliability of hypermnnesia for verbal materials has also been documented. For the present purposes, the important point to be gleaned from Erdelyi and Becker's account of hypermnnesia is that each successful recall of a target item was assumed to result in extensive marking of the search path that lead to the correct item. On any subsequent test(s), the subjects were presumed to locate the previously recalled items more quickly, thereby allowing additional time to attempt to retrieve other previously nonrecalled items and hence a hypermnnesic effect.

An alternative interpretation of hypermnnesia is the cumulative recall level hypothesis proposed by Roediger et al. (1982) and modified by Payne (1986). Roediger et al. (1982) argued that aspects of cumulative recall curves may provide insights concerning the necessary and sufficient conditions for producing hypermnnesia. A well-established empirical fact is that cumulative free recall curves typically show a negative correlation between the asymptotic level of recall and the rate of approaching that asymptote. This negative correlation means that, in general, with equal test durations across conditions, the conditions that produce high levels of asymptotic recall will be farther from asymptote at the end of Test 1 than will conditions that produce lower levels of asymptotic recall. The greater the difference between Test 1 recall levels and asymptotic recall levels, the more "room" there is for an improvement in recall levels, and hence the larger the hypermnnesic effect that can be expected.

As Roediger and Challis (1989) noted, the cumulative recall level hypothesis provides a reasonable functional description of hypermnnesia, and it is able to account for much of the available hypermnnesia data. The cumulative recall level hypothesis also identifies a number of factors that affect performance levels across repeated tests, including asymptotic cumulative recall level, rate of approaching the asymptote, performance levels on Test 1, and test length.

One assumption of the original cumulative recall level hypothesis that has proven to be invalid is the assumption of equivalent rates of intertest forgetting across conditions. Payne (1986) showed that when the cumulative recall levels for pictures and words were equated, pictures produced a lower rate of intertest forgetting and hence a larger hypermnnesic effect. Payne revised the cumulative recall level hypothesis and argued that two factors determine the magnitude of the hypermnnesic effect. The first factor is the difference between Test 1 net recall levels and asymptotic cumulative recall, with the magnitude of the hypermnnesic effect being correlated positively with this difference. The second factor is the extent to which retrieving an item increases that item's accessibility. Increasing item accessibility is assumed to decrease

intertest forgetting and increase the speed with which previously recalled items are retrieved on subsequent tests, thereby increasing the time that subjects have available to retrieve previously unrecalled items. Payne identified a number of factors (e.g., type of study materials, number of prior tests) that have been shown to affect the rate of intertest forgetting.

The third and most detailed interpretation of hypermnnesia offered to date is that of Raaijmakers and Shiffrin (1980), who proposed an explanation within the context of their Search of Associative Memory (SAM) model. To see how SAM accounts for hypermnnesia, we must first review briefly the manner in which recall is presumed to operate in the model.

An important feature of the SAM model is that retrieval from long-term memory is assumed to be a cue-dependent process (cf. Tulving, 1974). Probe cues are assembled in short-term memory, and these probe cues guide the sampling of items from long-term memory. In the SAM model, the cues that may guide retrieval are general contextual cues from the testing environment, other list items, and, in the case of categorized lists, category labels. The model also includes a number of mechanisms that determine when the retrieval process is halted; the stopping rules that are related to hypermnnesia are described below.

The SAM model accounts for hypermnnesia by positing two processes. The first process is called *incrementing*, which refers to the increase in the retrieved items' strength as well as the increase in the associative strength between the item and the cues present in short-term memory when the target item is recovered. There is considerable empirical evidence to support the notion of incrementing. For example, in repeated tests, subjects typically recall items faster on each subsequent test (see, e.g., Roediger & Payne, 1982). This is exactly what would be expected if each additional recall of an item served as a learning opportunity that allowed subjects to access the items more efficiently on subsequent tests. Note that this conclusion is in accord with Erdelyi and Becker's (1974) marking hypothesis and Payne's (1986) revised cumulative recall level hypothesis.

The second process with which SAM accounts for hypermnnesia comprises *alternate retrieval routes*. This refers to the notion that items not recovered when a given set of retrieval cues is present in short-term memory may be recovered later if at least one of the cues in the set changes. This notion is similar to Estes's (1955) stimulus sampling theory, in which element fluctuation accounts for the spontaneous recovery of conditioned responses following extinction.

According to Raaijmakers and Shiffrin (1980), the SAM model predicts the largest hypermnnesic effect when both of these processes are operational, and a smaller hypermnnesic effect when only one of the processes is active. Elimination of both processes results in no improvement in net recall across tests. To summarize, then, changes in the memory system that occur as a consequence of

recovering target items and variations in the specific cues used to guide retrieval are both implicated in the SAM account of hypermnnesia.

Although there are many differences between the accounts offered by Erdelyi and Becker (1974), Roediger et al. (1982), Payne (1986), and Raaijmakers and Shiffrin (1980), a number of common themes run through these interpretations, and each of these themes is related to the dynamics of memory retrieval. First, each view acknowledges the importance of the rate of item recoveries, either in terms of test-retest effects (i.e., faster recall of items across repeated tests) or the relation between asymptotic cumulative recall level and rate of approaching the asymptote. Second, with the exception of the original cumulative recall level hypothesis (Roediger et al., 1982), each interpretation stresses the increase in the likelihood that an item will be recovered on subsequent tests following an initial successful retrieval. This factor is directly related to intertest forgetting and indirectly related to item gains through the increase in functional retrieval interval afforded by the faster recovery of previously recalled items. The third common theme is that each interpretation assumes that underlying hypermnnesia there is a retrieval process that can be characterized as yielding diminishing returns over time. Erdelyi and Becker discussed searching through a memory network, whereas Raaijmakers and Shiffrin and Roediger et al. appeal to a stimulus sampling type of a notion. There thus seems to be some consensus regarding the importance of these three processes, although how these processes are instantiated in the various accounts differs considerably.

The next question to be addressed is how these retrieval dynamics factors might account for three major findings from the present experiments: the robust cued recall hypermnnesic effect; the similarity in the magnitude of the hypermnnesic effect in the cued recall and free recall conditions; and the increase in the access of categories and items within categories across tests (cf. Experiments 1 and 2).

As indicated above, there is considerable similarity among the three main theoretical amounts of hypermnnesia. For the purposes of exposition, we will couch our interpretation of the present findings within the framework of the SAM model. We will first consider how SAM can account for hypermnnesia in free recall and cued recall tests when categorized lists are employed.³

Raaijmakers and Shiffrin (1980) assume that when categorized lists are employed, the category labels may be used as cues to guide item recovery. On cued recall tests, these cues are explicitly presented during testing, and, according to SAM, subjects begin their memory search by using context cues and one of the category labels. The subjects continue to use each category label cue until there has been a specific number of failures to recover additional items with that category label. After this criterion number of retrieval failures occurs, the subject moves on to the next category label. Attempts at retrieving items continue until either the recall test ends or the total number of failures to recover new items exceeds a criterion.

The situation is somewhat different with free recall of categorized lists. Raaijmakers and Shiffrin (1980) proposed that subjects in free recall tests initially utilize context cues to sample items from long-term memory; when a list item is recovered, "the subject will generate (with a probability of 1.0) the category name of which that word is a member" (p. 238). Once a category member has been recovered and the category name has been generated, the category name continues to be used to guide retrieval in the same manner as in cued recall tests. Thus, according to SAM, cued recall and free recall of categorized lists differ only in that in free recall tests can the category labels be used as cues once a member of the category has been recovered. Put another way, recovery in cued recall tests is guided from the outset by the category cues, whereas in free recall tests, access to categories is governed by the retrieval of individual members of the categories. These initially recalled target items then serve, along with the category label, to guide in the recovery of other target items.

Because of the presumed similarities of the retrieval processes underlying free recall and cued recall, the same mechanisms that account for hypermnnesia in free recall should apply to cued recall, with the possible exception of category access. In both cued recall and free recall tests, there are two important search components, accessing categories and then searching for targets within the category. If one assumes that these searches are based on a sampling with replacement process that involves some fluctuation in the specific retrieval cues (i.e., context cues, item cues, category label cues) functioning over time, then both access to categories and recovery of items within categories ought to continue throughout the test period, with the most successful retrievals occurring early in the test period when the strongest cues will tend to control which items are recalled.

The fact that the hypermnnesic effect in the present experiments was of similar magnitude in free recall and cued recall makes sense if one assumes that the main difference between the two types of tests is initial access to the category. Cued recall will lead to direct access of the category members, because the category cues are presented in the test. In contrast, free recall will lead to category access when a category member is retrieved via context and/or item cues. The fact that the category access in the free recall test depends on retrieving a category member should increase the chances that there will be greater variability in the retrieval cues present in the short-term store for free recall than for cued recall. On the basis of Raaijmakers and Shiffrin's (1980) alternate retrieval route assumption, this in turn should facilitate the recovery of different items on each test and hence should facilitate higher levels of cumulative recall. This is exactly the pattern of results obtained with the related list condition in Experiment 1.

There is some converging empirical support for the view that hypermnnesia in free recall and cued recall reflect similar retrieval processes. First, Payne (1986, Experiment 4) reported that the main factor responsible for the hypermnnesic effect obtained with categorized lists was ac-

cess to additional categories across tests, and this was true for both free and cued recall tests. Payne also reported a small (but not significant) increase in the number of items per category recalled across tests. A similar pattern of results was obtained in Experiment 2 reported here, with the main difference being that the increase in the number of items recalled per category was significant in the present study. This difference between the two studies may be due to the fact that the recall periods used in the present study were 14% longer than those used by Payne. With additional time, subjects can continue to search longer, and this could result in additional items being recovered from the categories represented in the recall protocols.

Another set of findings that are consistent with the general retrieval dynamics view presented here is the failure to obtain hypermnnesia on recognition tests (e.g., Otani & Hodge, 1991; Payne & Roediger, 1987). If hypermnnesia depends on subjects' retrieving targets from memory on the basis of a set of retrieval cues that changes over time, one would not expect to find a hypermnnesic effect when the test is a straightforward recognition test and the target item is presented directly to the subjects.

The retrieval dynamics view also postulates that successful retrieval of target items serves to increase the strength of the target item as well as the associations between the target item and the retrieval cues that lead to the recovery of the target item. Roediger and his colleagues (Roediger & Challis, undated; cited in Roediger & Challis, 1989; Roediger & Payne, 1985) recently reported some interesting data that can be interpreted as reflecting just such an increase in strength as well as a fluctuation in the retrieval cues that guide recall.

Roediger and Payne (1985) were interested in whether changes in response criteria across repeated tests might affect the magnitude of the hypermnnesic effect. To address this possibility, they presented subjects with a list of items and then administered three recall tests. Different groups of subjects received free recall tests, forced recall tests, or "uninhibited free recall tests" in which the subjects were encouraged to report any items that came to mind. The free recall tests presumably allowed the subjects to vary their response criterion across tests, whereas the forced recall tests presumably controlled the subjects' response criteria by requiring a fixed number of responses on each test. The uninhibited free recall condition was intended to allow the subjects to adopt a very lax response criterion. If response criterion affects hypermnnesia, these three conditions should have differed in the size of the hypermnnesic effect they produced, with the forced recall condition producing the smallest effect and the uninhibited free recall condition showing the largest hypermnnesic effect. Roediger and Payne found that although all three groups showed a hypermnnesic effect, there was no difference in the magnitude of the effect across the three test types. Roediger and Payne argued that this finding poses problems for classical generate/recognize models of recall, which would predict that a more lenient response

criterion should result in more hits as well as more false alarms. The false alarm rates did vary across conditions as was expected, yet there was no difference in the magnitude of the hypermnnesic effect. Erdelyi, Finks and Feigin-Pfau (1989) have recently replicated and extended this finding and have shown that as long as subjects cannot easily guess the target items (on the basis of knowledge of the type of items in the list), there is in fact no advantage for free recall over forced recall.

An interesting twist to this finding was reported by Roediger and Challis (undated; cited in Roediger & Challis, 1989). Roediger and Challis presented subjects with 60 pictures and then tested them under either free recall instructions or forced recall instructions. For both types of recall tests, half of the subjects were given one test soon after the list was presented and then again 1 week later. The remaining subjects were only tested at the 1-week retention interval. The subjects who received the initial test and the delayed test recalled more items when given forced recall tests than when given free recall tests, and this pattern was obtained on both the immediate and delayed tests. Roediger and Challis also had the subjects rate their confidence that the items they recalled were from the study list. When the recall data were conditionalized on the subjects' recognition that the items were indeed target items, there was no difference between the free and forced recall conditions at either retention interval. Thus it seems that in these conditions forced recall subjects made more responses, but they did not recognize the additional items as being list items.

A different pattern of results was obtained with the groups that received only the single delayed test. Here there was a difference in the number of items recalled, with the forced recall subjects producing more correct responses than did the free recall subjects. However, this advantage of forced recall over free recall was also obtained when the data were conditionalized on subjects' recognition of the items. Although these findings appear to be reliable (Roediger and Challis have replicated this pattern of findings in a second experiment), Roediger and Challis offer no explanation for the effects of the initial test on the delayed test. We would argue that the retrieval dynamics view advocated here provides a reasonable explanation.

The retrieval dynamics view can potentially account for these findings through two mechanisms—the changes that occur when items are recalled, and the change in the strength of the context as a retrieval cue over the retention interval. According to the retrieval dynamics view, recalling items serves to increase the strength of the target items as well as the strength of the associations between the targets and the cues that resulted in the items' being retrieved. In agreement with this view, in Roediger and Challis's (undated; cited in Roediger & Challis, 1989) study, subjects who received the immediate test recalled more items on the delayed test than did subjects who received only the delayed test.

We also assume that context is a better cue at the immediate test than at the delayed test, either as a conse-

quence of a decrement in the context-item association strength or as a result of changes in the functional context between the two test sessions (e.g., the subject will be familiar with the testing environment on the second session but not the first). Note that there is some precedent for assuming that the effectiveness of context cues as retrieval cues will decrease across a retention interval; Mensink and Raaijmakers (1988) make similar assumptions in their contextual model of interference and forgetting.

If context serves as a strong cue on the immediate test, this will serve to make the free and forced recall conditions rather similar in terms of the retrieval cues that are used. Furthermore, at the time of the delayed test the situation is quite different for the single and repeated test conditions. In the repeated test condition, the item strength and item-to-item and item-to-context associations were increased on the initial test, and hence, on the delayed test, these associations and items would tend to be recovered. However, because of the change in context, the recovered items may appear less familiar to subjects and hence be given lower confidence ratings. For the single-test condition, at the time of the delayed test the interitem associations would tend to be the more effective cues than would the context-to-item associations. If interitem associations facilitate the retrieval of target items, then, because subjects in the forced recall condition were attempting to produce a set number of items, the items that they would produce would tend to be retrieved as a result of interitem associations.

Although this account is admittedly ad hoc, it does provide a reasonable interpretation for the effect of retention interval on recovery rates. It also provides a testable account in the sense that changes in context-to-item associations are assumed to be responsible for the observed differences between free and forced recall tests in the single- and repeated test conditions. Effective manipulations of context at encoding and/or retrieval should remove or reinstate the free versus forced recall test differences.

CONCLUSIONS

The retrieval dynamics view appears capable of accounting for a considerable number of findings from the hypermnnesia literature. On the basis of the available data, it seems clear that retrieval processes play a major role in the improvements in memory performance across tests and that additional research is needed to unravel the necessary and sufficient conditions for producing hypermnnesia. From a methodological point of view, cued recall tests are likely to be very informative in these efforts.⁴ Recognition tests do not typically lead to hypermnnesia, and free recall tests do not allow the researcher to exert much experimental control over retrieval processes. Cued recall tests also have the methodological advantage that they can be used to assess memory performance across the retention interval in situations in which free recall tests are rather unwieldy (e.g., prose memory). Another advantage of cued recall tests is that they may prove foren-

sically useful in memory improvement situations (e.g., eyewitness memory). Previous research (e.g., Scrivner & Safer, 1988; Tanenbaum, 1991) has shown that an eyewitness hypermnnesic effect can be obtained with free recall tests; on the basis of the present data, it appears that a similar effect may be obtainable with cued recall tests, and in this case, the investigator has more control over the cues used to aid recall of critical events.

One limitation of the extant cued recall hypermnnesia research is the reliance on categorically related cues and target items. Additional studies are needed to extend the range of retrieval cues used across repeated tests. According to the retrieval dynamics view proposed here, many different types of cues could result in a hypermnnesic effect.

REFERENCES

- ANASTASI, J. S., PAYNE, D. G., GOODMAN, L., LAMPKIN, C., & MARTINEZ, A. (1991, April). *Hypermnnesia, reminiscence, and interest forgetting: Encoding and retrieval factors*. Paper presented at the meeting of the Eastern Psychological Association, New York.
- ANDERSON, J. R., & BOWER, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review*, **79**, 97-123.
- BATTIG, W. F., & MONTAGUE, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monographs*, **80**(3, Pt. 2).
- ERDELYI, M. H. (1988). Hypermnnesia: The effect of hypnosis, fantasy, and concentration. In H. M. Pettinati (Ed.), *Hypnosis and memory* (pp. 64-94). New York: Guilford.
- ERDELYI, M. H., & BECKER, J. (1974). Hypermnnesia for pictures: Incremental memory for pictures but not words in multiple recall trials. *Cognitive Psychology*, **6**, 159-171.
- ERDELYI, M. H., FINKELSTEIN, S., HERRELL, N., MILLER, B., & THOMAS, J. (1976). Coding modality vs. input modality in hypermnnesia: Is a rose a rose a rose? *Cognition*, **4**, 311-319.
- ERDELYI, M. H., FINKS, J., & FEIGIN-PFAU, M. B. (1989). The effect of response bias on recall performance, with some observations on processing bias. *Journal of Experimental Psychology: General*, **118**, 245-254.
- ERDELYI, M. H., & KLEINBARD, J. (1978). Has Ebbinghaus decayed with time? The growth of recall (hypermnnesia) over days. *Journal of Experimental Psychology: Human Learning & Memory*, **4**, 275-289.
- ERDELYI, M. H., & STEIN, J. B. (1981). Recognition hypermnnesia: The growth of recognition memory (d') over time with repeated testing. *Cognition*, **9**, 23-33.
- ESTES, W. K. (1955). Statistical theory of spontaneous recovery and regression. *Psychological Review*, **62**, 145-154.
- HUNT, R. R., & EINSTEIN, G. O. (1981). Relational and item-specific information in memory. *Journal of Verbal Learning & Verbal Behavior*, **20**, 497-514.
- KLEIN, S. B., LOFTUS, J., KIHLMSTROM, J. F., & ASERON, R. (1989). Effects of item specific and relational information on hypermnnesic recall? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **15**, 1192-1197.
- MENSINK, G.-J., & RAAIJMAKERS, J. G. W. (1988). A model for interference and forgetting. *Psychological Review*, **95**, 434-455.
- OTANI, H., & HODGE, M. H. (1991). Does hypermnnesia occur in recognition and cued recall? *American Journal of Psychology*, **104**, 101-116.
- PAYNE, D. G. (1986). Hypermnnesia for pictures and words: Testing the recall level hypothesis. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **12**, 16-29.
- PAYNE, D. G. (1987). Hypermnnesia and reminiscence in recall: A historical and empirical review. *Psychological Bulletin*, **101**, 5-27.
- PAYNE, D. G., & ROEDIGER, H. L., III (1987). Hypermnnesia occurs

- in recall but not in recognition. *American Journal of Psychology*, **100**, 145-165.
- RAAIJMAKERS, J. G. W., & SHIFFRIN, R. M. (1980). SAM: A theory of probabilistic search of associative memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 14, pp. 207-262). New York: Academic Press.
- ROEDIGER, H. L., III, & CHALLIS, B. H. (1989). Hypermnnesia: Improvements in recall with repeated testing. In C. Z. Izawa (Ed.), *Current issues in cognitive processes: The Tulane Floweree symposium on cognition* (pp. 175-199). Hillsdale, NJ: Erlbaum.
- ROEDIGER, H. L., III, & PAYNE, D. G. (1982). Hypermnnesia: The role of repeated testing. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **8**, 66-72.
- ROEDIGER, H. L., III, & PAYNE, D. G. (1985). Recall criterion does not affect recall level or hypermnnesia: A puzzle for generate/recognize theories. *Memory & Cognition*, **13**, 1-7.
- ROEDIGER, H. L., III, PAYNE, D. G., GILLESPIE, G. L., & LEAN, D. S. (1982). Hypermnnesia as determined by level of recall. *Journal of Verbal Learning & Verbal Behavior*, **21**, 635-655.
- ROEDIGER, H. L., III, & THORPE, L. A. (1978). The role of recall time in producing hypermnnesia. *Memory & Cognition*, **6**, 296-305.
- SCRIVNER, E., & SAFER, M. A. (1988). Eyewitnesses show hypermnnesia for details about a violent event. *Journal of Applied Psychology*, **73**, 371-377.
- SHAW, G. A., & BEKERIAN, D. A. (1991). Hypermnnesia for high-imagery words: The effects of interpolated tasks. *Memory & Cognition*, **19**, 87-94.
- SMITH, S. M., & VELA, E. (1991). Incubated reminiscence effects. *Memory & Cognition*, **19**, 168-176.
- TANENBAUM, R. R. (1991). *The effects of memory strategies and repeated testing on eyewitness recall for high and low arousal events*. Unpublished doctoral dissertation, State University of New York, Binghamton.
- TULVING, E. (1974). Cue-dependent forgetting. *American Scientist*, **62**, 74-82.
- TULVING, E., & THOMSON, D. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, **80**, 352-373.
- WATKINS, O. C., & WATKINS, M. J. (1975). Buildup of proactive inhibition as a cue-overload effect. *Journal of Experimental Psychology: Human Learning & Memory*, **104**, 442-452.
- processes in recognition may be qualitatively different from those in recall. For example, generate/recognize theories (e.g., Anderson & Bower, 1972) view recall as involving an effortful process of generating potential target items followed by a decision process, whereas recognition is assumed to involve only a decision process.
2. The ANOVA assumption of homogeneity of variance was violated in the forgetting and recovery data from Experiment 3 because there was no forgetting in the cued recall condition. These data were thus analyzed further with separate dependent *t* tests for the free recall and cued recall conditions. To examine the component of recall $\times$ test type interaction, we calculated a difference score (recovery - forgetting) for each subject on each test pair and used these differences scores in a 2 (test type) $\times$ 2 (test pair) mixed factor ANOVA. The results from these analyses were generally consistent with the overall ANOVA findings. The difference scores indicated that item fluctuation rates decreased from Tests 1-2 to Tests 2-3 [$F(1,26) = 11.1$, $MS_e = 2.6$] and that there were somewhat greater fluctuations in the free recall condition than in the cued recall condition [$F(1,26) = 3.8$, $MS_e = 4.2$, $p = .06$]. There was also a marginally significant test type $\times$ test pair interaction [$F(1,26) = 3.4$, $MS_e = 2.6$, $p = .08$]. There was a significant difference between the free recall and cued recall conditions between Tests 1 and 2 ($p = .008$) but not between Tests 2 and 3 ($p = .70$). Finally, the difference between the rate of forgetting and item recovery was significant in the free recall condition between Tests 1 and 2 ($p \leq .001$) but not between Tests 2 and 3 ($p = .19$). In the cued recall condition, the forgetting and recovery rates were not significantly different for either test pair. Note, however, that this is likely due to the small sample size; no subjects forgot items between tests, and 9 of 14 subjects recovered items.
3. In our discussion, we assume that the target categorized lists are composed of common categories. However, the analysis can be extended easily to account for unrelated lists such as those used in Experiment 1. The main difference concerns (1) whether subjects can effectively use the category labels as retrieval cues, and (2) the likelihood that subjects in the free recall condition will generate the category labels when target items from the categories are recovered.
4. We thank Mathew Erdelyi for pointing out several of these methodological advantages of cued recall tests.

NOTES

1. Payne and Roediger (1987) did not assume that recognition tests involve no retrieval processes. Rather, they assumed that the retrieval

(Manuscript received August 5, 1991;
revision accepted for publication May 11, 1992.)