Hypermnesia occurs in recall but not in recognition

DAVID G. PAYNE
State University of New York at Binghamton

HENRY L. ROEDIGER III
Purdue University

Two experiments investigated the effect of encoding conditions and type of test (recall vs. recognition) on the phenomenon of hypermnesia (improved performance across repeated tests). Subjects in Experiment 1 studied a list of words using either imaginal or semantic elaboration strategies and then received three successive tests. Different groups of subjects received either free recall, four-alternative forced-choice recognition, or yes/no recognition tests. Reliable hypermnesia was found only in the recall conditions, with the recognition conditions showing either no change in performance levels across tests (forced-choice tests) or significant forgetting (yes/no tests). In Experiment 2, subjects studied a list of words, and encoding was manipulated using three orienting tasks. Once again, hypermnesia was found with the recall tests but not with the forced choice recognition tests. Finding hypermnesia in recall but not in recognition indicates that retrieval processes in recall play a major role in producing hypermnesia. Also, the finding that the magnitude of the recall hypermnesias increased with an increase in total cumulative recall levels across study conditions suggests that cumulative recall levels are an important factor in determining the presence or absence of recall hypermnesia.

The phenomenon of improved memory performance associated with repeated testing has attracted considerable attention in recent years. This finding was first reported in the modern experimental literature by Erdelyi and Becker (1974, Experiment 1) who presented subjects with a list of pictures and concrete nouns and then administered three 7-min recall tests. They found that the net recall level for pictures increased across tests, whereas the net recall level for words remained approximately constant across tests. Erdelyi and Becker termed the improvement in net recall levels they obtained with pictures "hypermnesia," to contrast it with amnesia, or forgetting.

Hypermnesia has now been obtained in a number of experiments employing a wide variety of materials and study conditions (see Payne,

AMERICAN JOURNAL OF PSYCHOLOGY
Summer 1987, Vol. 100, No. 2, pp. 145–165

1987, for a review). One consistent feature of these studies is the use of recall measures for assessing memory performance. Researchers have typically employed either standard free recall tests (e.g., Madigan, 1976; Madigan & Lawrence, 1980; Payne, 1986; Roediger & Payne, 1982, 1985) or a forced recall procedure (e.g., Belmore, 1981; Erdelyi & Becker, 1974; Erdelyi & Kleinbard, 1978; Popkin & Small, 1979; Roediger & Payne, 1985; Shapiro & Erdelyi, 1974) in which subjects are required to recall a specified (large) number of items, guessing if necessary.

Although some discrepancies have occurred across experiments using free versus forced recall tests, these two performance measures have generally yielded similar patterns of data. This conclusion regarding the consistency of recall hypermnesia receives support from an experiment reported by Roediger and Payne (1985) in which they directly compared performance levels across tests using several recall measures (free recall, forced recall, and an "uninhibited" free recall task in which subjects were to report anything that came to mind during the recall period). They found that whereas there were large differences in the intrusion rates for these three recall conditions (free < uninhibited < forced), the type of recall test had no effect upon either the magnitude of the hypermnesias obtained or the overall level of recall performance.

Although the finding that hypermnesia can be obtained with different recall measures demonstrates that the phenomenon is reliable, the fact that with a single exception (Erdelyi & Stein, 1981) researchers have used *only* recall tests to study hypermnesia raises an interesting question: Does hypermnesia represent a general improvement in memory performance that may be obtained in any repeated test paradigm or is hypermnesia a phenomenon limited strictly to recall? If hypermnesia can be obtained with other memory tasks, this would demonstrate its generality. On the other hand, if improvements across tests can only be obtained using recall measures, this would suggest that the processing demands associated with recalling items are critical in producing hypermnesia. Either way, knowledge of the extent to which hypermnesia depends upon the type of tests employed would increase understanding of the phenomenon and would also aid in efforts to provide a theoretical account of hypermnesia. For example, it has been suggested (e.g., Erdelyi & Becker, 1974; Raaijmakers & Shiffrin, 1980; Roediger & Thorpe, 1978) that the retrieval processes involved in recall play a major role in producing hypermnesia. If this is true, and if, as suggested by generate/recognize theories (e.g., Anderson & Bower, 1972; Kintsch, 1970; Shiffrin, 1970), recognition tests typically bypass an effortful retrieval stage, then we would not expect to find improvements across tests using repeated recognition tests.[1] If, however, hypermnesia represents a general improvement in performance associated with repeated testing, then we would expect to find hypermnesia with both recall and recognition tests.

To date only one attempt at inducing hypermnesia using tests other than recall measures has been reported. Erdelyi and Stein (1981) presented subjects with a series of single-frame cartoons consisting of a picture and a caption. For half of the study items, the caption and picture formed a meaningful scenario (configured items); for the remaining items, the caption and picture were unrelated (nonconfigured items). After studying these items subjects were given three yes/no recognition tests for their memory of either the pictures or the captions. Recognition hypermnesia, as indicated by increasing d' estimates of item accessibility across tests, was obtained only for the pictures from configured cartoons. None of the remaining three conditions (pictures in the nonconfigured pairings and captions in both pairings) showed any significant improvement across tests.

The Erdelyi and Stein (1981) data demonstrate that (under certain conditions) it is possible to induce recognition hypermnesia. However, Erdelyi and Stein employed study items (cartoons composed of both pictorial and verbal components) that are quite different from the items typically used to investigate recall hypermnesia (e.g., individual words or pictures of single objects). Hence, it is difficult to compare directly the recognition hypermnesia they obtained with the recall hypermnesia obtained in other studies. Furthermore, they did not include a condition in which subjects received several recall tests after studying the cartoons; hence, whether recall hypermnesia would have been obtained with the items and study conditions used in that experiment is an unanswered empirical question. The present research sought to obtain recognition hypermnesia under the same conditions that can be shown to produce recall hypermnesia.[2] Assuming an affirmative answer, a related question was whether recognition hypermnesia is affected by the same factors that influence recall hypermnesia.

A second issue addressed in this study concerns the appropriate theoretical interpretation of the hypermnesia phenomenon. Two interpretations that have attracted attention recently are the imagery hypothesis (Erdelyi & Becker, 1974) and the cumulative recall level hypothesis (Roediger, Payne, Gillespie, & Lean, 1982). Erdelyi and Becker (1974) proposed the imagery hypothesis to account for their finding of hypermnesia for pictures but not for words. According to this view, the picture-word difference in hypermnesia is due to differences in the formats in which these items are represented in memory (i.e., an imaginal vs. a verbal/linguistic code). Using a generate/

recognize model of recall, Erdelyi and Becker suggested that there may be differences in the sensitivity of the recognition stage for memory traces arising from pictures as opposed to words (pictures > words). Some support for this assumption comes from the well-documented difference in performance on recognition memory tests for pictorial versus verbal stimuli (e.g., Standing, Conezio, & Haber, 1970). This difference in recognizability of pictorial versus verbal items is assumed to impact the hypermnesia phenomenon by affecting both the rate of item recovery and the intertest forgetting rate (i.e., less intertest forgetting with pictures than words).

The imagery hypothesis is consistent with several empirical observations. First, several studies reported early in this century obtained hypermnesia using items such as easily imageable poetry (e.g., Ballard, 1913) or concrete objects (e.g., Nicolai, 1922, cited in Tulving, 1968) that presumably could be coded in an imagistic format. Second, many studies failed to obtain hypermnesia using verbal materials such as word lists (e.g., Donaldson, 1971; Hogan & Kintsch, 1971; Tulving, 1967). Third, Erdelyi, Finkelstein, Herrell, Miller, and Thomas (1976) demonstrated that a sizable hypermnesic effect can be obtained with words when subjects are encouraged to employ imaginal encoding strategies. Finally, even when hypermnesia for words is obtained without instructing subjects to engage imaginal encoding strategies (e.g., Belmore, 1981, Experiment 1; Dragone, Brown, Krane, & Krane, 1980; Payne, 1986; Roediger & Thorpe, 1978, Experiments 1 & 2), the magnitude of this hypermnesic effect is generally less than that obtained with pictures (e.g., Payne, 1986). Thus the imagery hypothesis provides a reasonable account of a number of different findings in the experimental literature (see Erdelyi, 1984, for a more complete account of the imagery hypothesis and data consistent with that view).

Roediger et al. (1982) proposed an alternative interpretation of hypermnesia when they noted that in many hypermnesia experiments, variables that raise recall levels (e.g., type of study materials, encoding instructions) also generally increase the magnitude of the hypermnesic effect (e.g., Belmore, 1981, Experiments 1 & 2; Erdelyi & Becker, 1974; Erdelyi et al., 1976; Rodeiger & Thorpe, 1978, Experiments 1 & 2). Furthermore, it is well documented from studies examining cumulative recall (e.g., Bousfield & Sedgewick, 1944) that conditions producing (relatively) low recall levels generally reach asymptotic performance levels more quickly than do conditions yielding higher asymptotic cumulative recall levels. Hence, after the first test, there is more "room" for improvement in net recall levels on subsequent tests (i.e., hypermnesia) in conditions with high cumulative recall level asymptotes than there is in conditions with lower asymptotic performance levels. Based on these observations, Roediger et al. argued that hypermnesia is directly related to the level of asymptotic cumulative recall within the hypermnesia experiment, regardless of the factor(s) responsible for producing the observed recall levels. According to this view, any factor that raises cumulative recall levels should also increase the likelihood of obtaining a hypermnesic effect. (See Roediger, 1982, for a review of data supporting this general approach.)

Although the cumulative recall level hypothesis is consistent with much of the hypermnesia literature, Payne (1986) has recently presented data that seem to be an exception. Payne demonstrated (Experiments 1 & 2) that when the cumulative recall levels for words were raised above those obtained with pictures (by varying the number of list presentations or the rate of item presentation), pictures still produced larger hypermnesic effects than did words. Further analyses showed that picture superiority in hypermnesia experiments occurs because pictures produce less intertest forgetting than do words. This finding highlights the need to consider intertest forgetting rates (as well as item recovery, or reminiscence) in providing a general account of the hypermnesia phenomenon.

EXPERIMENT 1

Experiment 1 was designed to assess the role of the type of test in producing hypermnesia and to provide another test between the imagery and cumulative recall level interpretations of hypermnesia. With regard to the second issue, we asked three straightforward questions. First, can hypermnesia be obtained with words when subjects are not instructed to form images of the words' referent but rather are instructed to process the meaning of the target item? We assume, along with previous researchers (e.g., Erdelyi & Becker, 1974), that subjects are more likely to store items in an imaginal format when they are explicitly instructed to use imaginal coding than when they are instructed to use other, nonimaginal, encoding strategies. Second, using word lists, do imaginal encoding instructions enhance the hypermnesic effect above that observed with semantic elaboration study instructions that often produce recall levels equal to or greater than those observed with imaginal instructions (cf. Roediger & Payne, 1985)? If both types of study instructions produce hypermnesia, then the third question we can ask is whether the imaginal or semantic elaboration study instructions produce differing amounts of hypermnesia.

To determine the effect of the type of test on hypermnesia, we presented subjects with a list of common English words and then gave them either three free recall tests, three 4-alternative forced-choice (4AFC) recognition tests, or three yes/no recognition tests. Two types of recognition tests were employed so that we could examine performance levels across tests in which subjects' response criteria were assumed either to be held constant (forced-choice tests) or allowed to vary (yes/no tests).

The data from the recall conditions provide answers to the three questions described previously concerning the imagery and recall level hypotheses. By comparing subjects performance levels across the recall and recognition tests, we will obtain evidence relevant to the role of the test type in producing hypermnesia. Finally, the data from both the recall and the 4AFC recognition test conditions provide information concerning the relation between intertest forgetting and hypermnesia. Payne (1986) demonstrated that intertest forgetting rates are affected by the type of study materials employed (pictures < words). In the present experiment we asked whether the type of study instructions (imaginal vs. semantic elaboration) also affect intertest forgetting across the repeated recall tests. Analysis of the intertest forgetting rates in the 4AFC recognition condition will indicate the extent to which performance levels across repeated recognition tests are affected by the "recovery" of old items (i.e., an increase in the number of target items correctly recognized) and the failure to recognize previously recognized items.

METHOD

Subjects

Participants were 108 introductory psychology students at Purdue University who served in partial fulfillment of a course requirement. Subjects were assigned to conditions based on their order of arrival at the experiment and were tested in groups of 8 or fewer.

Design

A 2 × 3 × 3 mixed factorial design was employed. Six between-subjects groups (n = 18) were formed by factorially combining two types of study instructions (imagery vs. semantic elaboration) with three types of memory tests (free recall, 4AFC, and yes/no recognition). The only within-subjects factor was tests (1, 2, 3).

Materials

A pool of 400 common English nouns was selected. Target items were 100 randomly chosen words that were typed and prepared as slides. The remaining 300 words served as distractors in the recognition tests. The mean imagery rating (Paivio, Yuille, & Madigan, 1968) and normative frequency of occurrence (Kucera & Francis, 1967) of the target items for which these data were available was 6.5 (n = 26) and 49.9 (n = 94), respectively. Subjects in the recognition test conditions received three alternate forms of the appropriate test, with 3 subjects assigned to each of the six possible orderings of these three test forms. Each of the three yes/no recognition tests contained the 400 items in a different serial order. For the alternate forms of the 4AFC tests, the targets and lures were randomly reordered on each test. Across tests no target item appeared with the same distractor, and the placement of targets within a quartet (i.e., first, second, etc.) was also varied.

Procedure

Subjects were told that they would be presented with 100 common English words via slide projector and that their task was to attempt to memorize these items using a study strategy that the experimenter would provide. Subjects in the imagery condition were told that as each word was presented they should try to form an image of the item's referent, making a special effort to form a "clear and vivid" image. Subjects in the semantic elaboration condition were instructed to try to think of both the meaning of each word as well as "things they normally associate with" that item. Subjects in both conditions were told that they would be given instructions for the testing phase of the experiment after the list was presented. They were not informed of the type of memory test they would receive or that more than one test would be given.

The study list was then presented at a 5-s rate. Immediately after presentation of the list, subjects were given brief (1–2-min) instructions appropriate for their test condition. Subjects in the free recall condition received standard instructions to recall as many words as possible in any order, with the exception that they were also told that every minute during the test period they would be asked to draw a line beneath the last item they had written. Subjects were asked to draw a line after each minute of the 12-min tests. Subjects in the 4AFC condition were told that their test contained 100 rows of four words each and that one of these four items had appeared in the slides. They were told to circle the one word on each line that they thought had appeared in the study list. The instructions for the yes/no recognition test indicated that the test contained the 100 target items along with 300 distractors. Subjects were told that they were to check Y (for yes) if they thought the word had appeared in the study list or check N (for no) if they felt it was not an item from the study list.

Test 1 began immediately after these instructions were read and subjects' questions were answered. This test lasted 12 min, with the experimenter giving a warning when there was 2 min remaining. (For the two recognition test conditions a 1-min grace period was given if subjects had not completed Test 1 within 12 min. No grace period was given on Tests 2 and 3, each of which lasted 12 min.)

At the end of Test 1, subjects were told that their task on the (unexpected) second test was the same as on Test 1, namely to recall/recognize as many items from the original list as possible. Subjects in all conditions were encouraged to try to improve their performance on Test 2 relative to Test 1. Subjects were warned, however, to avoid guessing wildly. Test 2 began as soon as these brief instructions were read. Test 3 was similar to Test 2 except that subjects were told that Test 3 would be the last test.

RESULTS AND DISCUSSION

For the free recall and the 4AFC recognition test conditions, the number of items correctly recalled or the proportion of items correctly recognized were used as the dependent measures. Performance in the yes/no recognition condition was assessed by computing d' values (Hochhaus, 1972) from the subjects' hit and false alarm rates. The recall protocols were also used to measure cumulative recall levels. In computing these measures, subjects were given credit for an item the first time it was recalled and any subsequent occurrences of that item were disregarded. This measure provides us with the total (i.e., cumulative) number of items that were recalled (at least once) across the entire test period. Finally, intertest forgetting and recovery rates were established for subjects in the free recall and 4AFC recognition test conditions. Intertest forgetting refers to not recalling (recognizing) an item on Test I + 1 that was recalled (recognized) on Test I. Item recovery refers to the opposite situation in which an item is recalled (recognized) on Test I + 1 that was not recalled (recognized) on Test I. (The data from the yes/no recognition condition were not analyzed in this fashion because any such changes in "forgetting" or "recovery" could reflect changes in the subjects' response criteria and not changes in the accessibility of the individual items.)

Performance in the three test conditions is shown in Table 1 and was analyzed using separate 2 (Encoding Instructions: imagery vs. semantic elaboration) × 3 (Test: 1, 2, 3) mixed-factor analysis of variance (ANOVA) tests, one for each test condition (free recall, 4AFC, yes/no). As expected, the free recall conditions produced a reliable hypermnesic effect, $F(2, 68) = 22.645$, $MS_e = 8.58$, with each condition showing consistent improvements in mean net recall levels across the three tests. (All effects are reliable at the .05 level of confidence unless otherwise specified.) Another interesting finding was that the semantic elaboration condition produced higher net recall levels than did the imagery condition, $F(1, 34) = 6.41$, $MS_e = 309.8$. The same superiority of the semantic elaboration over the imagery condition was also observed in the total cumulative recall levels (see Figure 1). Finally, consistent with the predictions of the cumulative recall level hypothesis, the semantic elaboration study condition produced a larger hypermnesic effect (i.e., Test 3 − Test 1) than did the imagery condition (6.2 vs. 3.0 items, respectively). This conclusion should be viewed with some caution, however, as the Encoding Condition × Tests interaction only approached the criterion for statistical significance, $F(2, 68) = 2.82$, $MS_e = 8.58$, $p = .065$.

The results from the free recall conditions allow us to address several issues raised in the Introduction. First, hypermnesia was obtained in

Table 1. Experiment 1: Mean performance levels on the three tests and changes in performance levels across tests for all six conditions

Encoding condition	Test 1	Test 2	Test 3	Test 3 − Test 1	% change
Free recall: Mean proportion of items correctly recalled					
Semantic elaboration	.39	.42	.45	+.06	+15.8
Imagery	.30	.32	.33	+.03	+ 9.9
4AFC recognition: Mean proportion of items correctly recognized					
Semantic elaboration	.86	.86	.86	.00	.00
Imagery	.88	.89	.88	.00	.00
Yes/no recognition: Mean d' values					
Semantic elaboration	1.83	1.72	1.53	−.30	−16.4
Imagery	2.98	2.78	2.65	−.33	−11.1

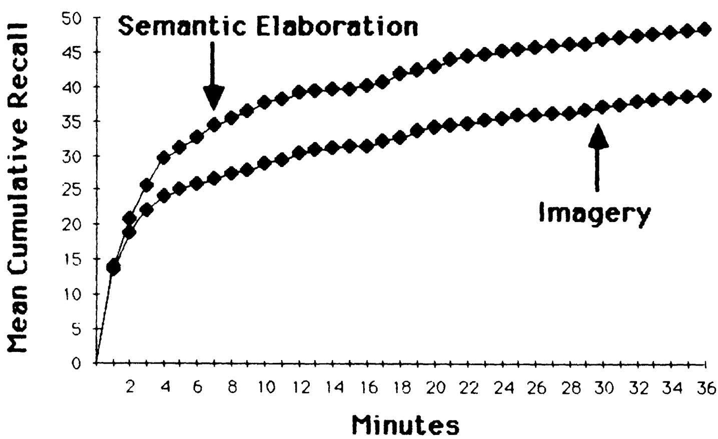

Figure 1. Mean cumulative recall curves across the entire 36-min test period for the imagery and semantic elaboration conditions in Experiment 1

the absence of explicit imaginal encoding instructions. Second, there was no evidence that imaginal coding produced greater hypermnesia than meaningful semantic processing. Finally, there was some evidence that the semantic elaboration encoding condition produced greater hypermnesia than did the imaginal encoding condition. Taken together, then, these and other data (e.g., Roediger et al., 1982) indicate that imaginal coding is not a necessary condition for producing recall hypermnesia.

Given that the materials and study conditions employed in Experiment 1 produced recall hypermnesia, the next question asked was whether these same materials and conditions were sufficient to induce recognition hypermnesia. The answer to this question appears to be no, because recognition performance levels either remained constant across tests (4AFC: $F < 1.0$) or showed a significant decline, thus indicating forgetting as opposed to hypermnesia; yes/no: $F(2, 68) = 24.83$, $MS_e = .043$. The null result observed in forced choice recognition may be due to a ceiling effect, a possibility examined in Experiment 2. The only other significant effect was a main effect of encoding condition obtained in the yes/no recognition test condition, $F(1, 34) = 23.42$, $MS_e = 1.246$. The imagery instructions produced higher yes/no recognition performance levels than the semantic elaboration instructions.

In a further analysis we examined intertest forgetting and recovery for the free recall and 4AFC recognition test conditions. As shown in Table 2, recovery exceeded forgetting in every possible case in the free recall conditions, $F(1, 34) = 33.58$, $MS_e = 5.36$, thereby producing the hypermnesic effect observed in the net recall data (see Table 1). Furthermore, replicating previous studies (e.g., Payne, 1986; Roediger & Payne, 1982), there was a larger fluctuation in the specific items recalled between Tests 1 and 2 than between Tests 2 and 3, $F(1, 34) = 4.44$, $MS_e = 3.75$. None of the other main effects or interactions were significant in the analyses of the free recall data. For the 4AFC condition, there were no significant main effects or interactions, thus indicating that the small amount of item fluctuation that was observed in this test condition was not related systematically to either the encoding operations (i.e., imagery vs. semantic elaboration) or the specific tests (i.e., Tests 1 and 2 vs. Tests 2 and 3).

The results of Experiment 1 are clear in showing improvement across tests in free recall, but not on either of two types of recognition tests. However, the effect of study strategies across tests was puzzling: The semantic strategy produced better performance in free recall than the imagery strategy, the reverse occurred in yes/no recognition, and no difference appeared in the forced-choice recognition test condition. We have no explanation for this pattern, nor would we be inclined to invent one until the result is replicated. However, this unusual result does not compromise our primary conclusion that hypermnesia occurs in recall, but not recognition, under the same study conditions.

EXPERIMENT 2

Erdelyi and Stein (1981) obtained significant recognition hypermnesia for pictures from configured cartoons. However, in Experiment 1 we found no improvement in memory performance for words tested with either yes/no or forced-choice recognition tests, despite the fact that the materials and study conditions we employed were sufficient to produce hypermnesia in free recall. Experiment 2 was designed to examine the generality of the recall versus recognition difference in hypermnesia. Toward this end, Experiment 2 was designed to increase the likelihood of obtaining recognition hypermnesia. Several lines of evidence from Experiment 1 suggested that (using materials similar to those used in Experiment 1) forced-choice recognition tests are more likely to yield hypermnesia than are yes/no recognition tests. First, the yes/no recognition conditions in Experiment 1 produced significant forgetting, as opposed to hypermnesia. Second, although there were no improvements in performance levels across tests in the

Table 2. Experiment 1: Mean number of items forgotten and recovered between successive tests in the free recall and 4AFC recognition test conditions

Encoding condition	Performance measure	Test trials 1–2	Test trials 2–3
	Free recall		
Semantic elaboration	Forgetting	3.66	2.11
	Recovery	6.27	5.44
Imagery	Forgetting	3.72	3.27
	Recovery	4.94	5.05
	4AFC recognition		
Semantic elaboration	Forgetting	4.66	4.88
	Recovery	4.50	5.27
Imagery	Forgetting	3.88	4.61
	Recovery	4.38	4.38

forced-choice conditions, there was no decrease in performance levels either. Third, intertest recovery rates in this condition were reasonably high. Finally, in the 4AFC condition performance levels on Test 1 were fairly high, and thus ceiling effects may have prevented us from detecting possible improvements across tests.

To avoid possible ceiling effects in the forced-choice recognition test condition, we incorporated three methodological changes in Experiment 2. First, we increased the number of distractors per target item from three to five. Second, to make the discrimination task more difficult (and hence to lower performance), we used semantically related targets and distractors. Third, we included a levels-of-processing manipulation (Craik & Tulving, 1975) so that we could examine performance across a range of performance levels. These three aspects of Experiment 2 should allow us to determine whether the failure to obtain recognition hypermnesia with the forced-choice test conditions used in Experiment 1 was simply due to a ceiling effect.

With regard to the recall conditions, the levels-of-processing manipulation provides a further test of the adequacy of the cumulative recall level hypothesis. In Experiment 1 the semantic elaboration condition produced the highest cumulative recall levels and, consistent with the cumulative recall level hypothesis, also produced the largest improvement in net recall levels across the three tests. However, the Encoding Conditions × Tests interaction was only marginally significant. This latter finding may have occurred because the semantic elaboration and the imagery condition both produced relatively high recall levels. The study conditions used in Experiment 2 were intended to produce a broader range of overall cumulative recall levels. The levels-of-processing manipulation also provides a conceptual replication of an experiment reported by Roediger et al. (1982, Experiment 1) in which recall levels were varied using a levels-of-processing manipulation. Results showed that the magnitude of the hypermnesic effect was related directly to the observed cumulative recall levels. The primary interest here is in seeing if similar effects will be observed in a forced-choice recognition test.

Subjects in Experiment 2 studied a list of words under three different encoding conditions. As the list items were presented, subjects rated them on one of three dimensions: graphemic, semantic, or imaginal. After studying the target items, subjects received either three recall tests or three recognition tests. It was expected that the three encoding conditions would produce a broad range of performance levels in both the recall and recognition test conditions, with the graphemic orienting condition producing low recognition rates and the semantic and imaginal encoding conditions producing fairly high performance levels. Thus the two primary goals of Experiment 2 were (a) to attempt once more to induce recognition hypermnesia using conditions that should increase the chances of obtaining such an effect, and (b) to determine whether the magnitude of recall hypermnesia is related to the observed level of recall, as suggested by the cumulative recall level hypothesis.

METHOD

Subjects

Participants were 144 volunteers selected from the same source as in Experiment 1. Subjects were assigned to conditions based on their order of arrival at the experiment and were tested in groups of 10 or fewer.

Design

Each of 12 independent groups of subjects studied the same list of 99 target items in the same serial order. Encoding condition was varied as a within-subjects factor, with each subject rating one third of the target items in a block on a graphemic dimension (number of letters), another third on an imaginal dimension (ease of imageability), and the final third on a semantic dimension (pleasantness). Order of encoding conditions was counterbalanced across subjects, thus producing six orders. After an initial study phase, half of the subjects from each encoding group received three free recall tests while the remaining subjects received three 6-alternative forced-choice (6AFC) recognition tests. The design was thus a 3 × 2 × 3 mixed-factor design with Encoding Condition (letters, imagery, pleasantness) and Test (1, 2, 3) being within-subjects factors and Type of Test (recall vs. recognition) being manipulated between-subjects.

Materials

The pool of 594 common English words used in the experiment consisted of 99 sextuplets, each containing six semantically related items (e.g., weak, proud, villain, captain, strong, coward). These were originally constructed by Schacter (1983) for use in feeling-of-knowing experiments. Within each sextuplet, one item was randomly selected as the target with the remaining five items serving as distractors on the 6AFC recognition tests. Each target item was typed in lowercase letters, prepared as a slide, and randomly assigned to a serial input position (i.e., 1–99). This study list was then divided into three blocks of 33 items each (i.e., Items 1–33, 34–66, 67–99) for manipulating encoding conditions.

Subjects in the 6AFC condition received three different forms of the recognition tests, with order of test form counterbalanced across subjects. The three forms of the 6AFC recognition test were constructed such that each target item appeared with the same five (semantically related) distractors on each test, but the order of items within the sextuplet (i.e., 1–6) was varied across tests as was the serial position of sextuplets within each test (i.e., 1–99).

Procedure

Subjects were told that they would see a list of 99 common English words and that their memory for these items would be tested following list presentation. No mention was made of either the type of test subjects would receive or that they would be given more than one test. Subjects were also told that each successive group of 33 items would be rated on a different dimension. These ratings were recorded in booklets containing three sheets of 33 lines each, with either a 1 to 9 rating scale (imagery and pleasantness conditions) or a 2 to 10 scale (number of letters condition). Subjects were told that before each group of 33 items was presented, the experimenter would describe the rating dimension for that set of items. Words were presented at a 5-s rate with a short break between successive thirds of the list, during which time the experimenter read the instructions appropriate for the next rating task. These instructions asked the subjects to rate either the number of letters in each word, how pleasant each word was (1 = very unpleasant), or how easy it was to form an image of the words' referent (1 = very difficult to form an image).

After presentation of the list, the experimenter read a brief set of instructions appropriate to the condition being tested. These test instructions were identical to those used in Experiment 1, except that subjects in the recognition test condition were told that each line of the test contained six items: one target and five distractors. Subjects in both conditions were given 12 min to complete the test, with warnings being given when 2 min and 1 min remained. At the end of Test 1, subjects turned to the next (blank) sheet in their test booklets and were given instructions for the second test. Subjects were encouraged to try to improve their performance above that of Test 1. Subjects were told that Test 3 would be the last test.

RESULTS AND DISCUSSION

Presented in Table 3 are the mean proportion of items correctly recalled or recognized across the three tests for each of the three encoding conditions and the percentage change across the three tests. Consistent with Experiment 1, reliable hypermnesia was obtained in the recall test condition, $F(2, 142) = 70.7$, $MS_e = .003$, but there was no change in performance levels across the three recognition tests, $F(2, 142) = 1.7$, $MS_e = .002$, $p > .05$. As expected, varying the initial orienting task affected performance levels in both the recall test condition, $F(2, 142) = 80.9$, $MS_e = .02$, and the recognition test condition, $F(2, 142) = 312.3$, $MS_e = .03$. The levels of processing manipulation also affected the observed total cumulative recall levels, with the mean number of unique items recalled in the letters, imagery, and pleasantness conditions being 7.26, 13.52, and 14.66, respectively, $F(2, 142) = 92.5$, $MS_e = 12.37$. Consistent with the cumulative recall level hypothesis, the magnitude of the recall hypermnesia was related directly to the overall recall performance level, as indicated by the significant Encoding Condition × Tests interaction, $F(4, 284) = 10.4$, $MS_e = .002$, thus replicating Roediger et al. (1982, Experiment 1).

GENERAL DISCUSSION

The primary goal of these experiments was to determine whether recognition hypermnesia could be obtained using study conditions known to produce recall hypermnesia. Our results suggest that the answer to this question is no: In two experiments we obtained hypermnesia with words using repeated recall tests, but not repeated recognition tests. This suggests that the recognition hypermnesia reported by Erdelyi and Stein (1981) may not generalize to other materials and test conditions. Numerous differences exist between the Erdelyi and Stein (1981) study and the present experiments (e.g., type of target items, initial encoding conditions, nature of the recognition distractors, etc.). Based on the extant data, it is impossible to determine which (if any) of these factors is responsible for producing the differences in results between these studies. However, the conclusion that recognition hypermnesia is less general than hypermnesia in free recall seems likely, because in both our experiments we found hypermnesia in recall but not in recognition.

Why should hypermnesia occur in free recall but not in recognition? Although our data do not permit a definitive answer to this question,

Table 3. Experiment 2: Mean performance levels on the three tests and changes in performance levels across tests for all six conditions

Encoding condition	Test 1	Test 2	Test 3	Test 3 − Test 1	% change
Free recall: Mean proportion of items correctly recalled					
Number of letters	.16	.16	.18	+.02	+12.5
Imagery	.27	.30	.35	+.08	+29.6
Pleasantness	.31	.34	.39	+.08	+25.8
6AFC recognition: Mean proportion of items correctly recognized					
Number of letters	.52	.49	.50	−.02	−3.8
Imagery	.83	.83	.84	+.01	+1.2
Pleasantness	.86	.85	.86	.00	0.0

we would offer the following interpretation. Free recall likely involves a recursive retrieval process such that properly motivated subjects will continually "sample" information from memory over long periods of time with continually diminishing returns (Indow & Togano, 1970; Raaijmakers & Shiffrin, 1981). The outcome of this process can be seen in the gradually increasing cumulative recall curves in Figure 1. The mechanism responsible for the item recoveries across the recall period may be shifting of the internal/covert retrieval cues subjects use to mediate recall (Erdelyi & Kleinbard, 1978; Payne, 1986; Roediger & Thorpe, 1978; Tulving, 1974). This mechanism can account for cumulative recall curves such as those shown in Figure 1, but it cannot account for why performance levels on each individual test increase. However, it has been shown (e.g., Ballard, 1913; Roediger & Payne, 1982) that when subjects are given repeated free recall tests, their rate of recalling items increases across each successive test. Thus performance on later tests benefits from more readily recalling items that were already recalled on previous tests. The more efficient recall of old items thus permits more time on later tests for recall of additional new items, at least to the asymptotic limit. Factors that increase asymptotic recall, such as semantic coding in Experiments 1 and 2 here, should also increase hypermnesia, as we found. These two factors—improved retrieval of previously recalled items and continued retrieval of new ones due to shifts in cues—provide a tentative account of hypermnesia (for more details see Payne, 1986; Roediger, 1982).

The situation in recognition tests is quite different. Rather than using internal cues, subjects are typically supplied with strong "copy cues" on their tests. They examine each and note whether it seems to represent a list member. Although some sort of retrieval process doubtless operates in recognition, its nature would seem quite different from that in free recall. In particular, in recognition a strong cue is given for all items on the first test, and thus item recoveries when the same cue is given later are less likely. In addition, because the old items are embedded in so many other distractors on recognition tests, the benefit of a first correct recognition on a later recognition test is likely to be less than is an initial recall on a later recall test. In short, neither factor that operates in recall to produce hypermnesia is as potent in a recognition test in which complete "copy cues" are provided on each test. The same would likely be true of cued recall tests, too, because typically powerful cues are given for studied items and those same cues would be repeated on later tests. Thus, according to this interpretation, hypermnesia should occur in either (a) free recall situations, in which subjects cannot generate internal cues for all items on a first test, or (b) other types of test that impose large retrieval demands.

The foregoing analysis is consistent with the views of Erdelyi and Stein (1981), who suggested that "For recognition hypermnesia to be obtained . . . retrieval search must be a non-trivial component of the recognition task" (p. 30). Note that in the Erdelyi and Stein experiment, subjects were not given a copy of the initially studied items (i.e., the cartoon and the caption), but rather they were given one component of these items (either the cartoon or the caption) and were asked to determine whether this component corresponded to one of the (complete) target items from the study list. This decision process presumably involves attempting to retrieve or redintegrate the intact target item (cartoon plus caption) and hence would require retrieval of the component of this item that was not presented on the recognition tests. It seems likely that this retrieval process would be more likely to succeed when the cartoon and caption "match" (i.e., the configured items). This (along with differences in the recognizability of pictorial and verbal items) could account for why recognition hypermnesia was obtained with pictures from configured targets but not from nonconfigured items.

Finally, the difference in performance between subjects given free-choice (yes/no) and forced-choice tests in Experiment 1 is worth noting. Performance declined across successive tests for free-choice subjects, but stayed at the same high level for forced-choice subjects. In the yes/no tests, subjects examined all 400 items in a different order each time, attempting to recognize the 100 studied items. Since distractors on the first test doubtless began to look familiar on the second and third tests, such items would be more likely to elicit false alarms. Indeed, the false alarm rates across Tests 1, 2, and 3 were 4.2%, 7.1%, and 9.1%, respectively. Of course, distractors also became increasingly familiar across the forced-choice tests, but in this case subjects had only to select the most familiar item from the set. Thus, because each set contained a list item, its greater familiarity promoted good performance that did not decline across tests, unlike the free-choice case. The upshot of this is that forced-choice recognition tests are probably to be preferred over free-choice tests for repeated assessments of retention. In free choice, the repeated presentation of the distractor items on the recognition tests should increase the overall familiarity of these items. To the extent that subjects are unable to determine accurately whether the increase in familiarity is due to presentation during the study phase as opposed to the test phase, this would decrease their sensitivity in terms of being able to discriminate items that are familiar because they appeared in the study list as compared to items that are familiar because they appeared in the previous tests (as distractors).

Notes

Preparation of this manuscript was supported in part by a New York State/ United University Professors New Faculty Development Award to David G. Payne. We thank Harley A. Bernbach for help in developing this research. Requests for offprints should be sent to David G. Payne, Department of Psychology, State University of New York at Binghamton, Binghamton, NY 13901. Received for publication March 18, 1986; revision received July 27, 1986.

1. Of course, we do not assume that recognition involves no retrieval process; rather we assume that the retrieval processes in recognition may be qualitatively different from those in recall.

2. We should point out that there is a sizable literature that deals with the effects of the types of test in repeated test paradigms (e.g., Birnbaum & Eichner, 1971; Bregman & Wiener, 1970; Hogan & Kintsch, 1971; McDaniel & Masson, 1985; Rosner, 1970; Thompson, Wenger, & Bartling, 1978; Wenger, Thompson, & Bartling, 1980). However, these studies are not really germane to the question at hand, because hypermnesia was seldom obtained in these studies even when recall tests were employed. The failure to induce recall hypermnesia in these studies is probably due to the use of very short test periods, because it has been shown (e.g., Roediger & Payne, 1981) that for the type and number of study items typically employed, differences in performance across tests are obtained only when the test periods last at least several minutes. To ensure that recall hypermnesia would be obtained, we employed fairly long test periods.

References

Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review, 79,* 97–123.

Ballard, P. B. (1913). Oblivescence and reminiscence. *British Journal of Psychology Monograph Supplements, 1,* 1–82.

Belmore, S. M. (1981). Imagery and semantic elaboration in hypermnesia for words. *Journal of Experimental Psychology: Human Learning and Memory, 7,* 191–203.

Birnbaum, I. M., & Eichner, J. T. (1971). Study versus test trials and long-term retention in free-recall learning. *Journal of Verbal Learning and Verbal Behavior, 10,* 516–521.

Bousfield, W. A., & Sedgewick, C. H. W. (1944). An analysis of sequences in restricted associative responses. *Journal of General Psychology, 30,* 149–165.

Bregman, A. S., & Wiener, J. R. (1970). Effects of test trials in paired-associate and free-recall learning. *Journal of Verbal Learning and Verbal Behavior, 9,* 689–698.

Craig, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104,* 268–294.

Donaldson, W. (1971). Output effects in multitrial free recall. *Journal of Verbal Learning and Verbal Behavior, 10,* 577–585.

Dragone, G. A., Brown, M., Krane, A. R., & Krane, R. V. (1980). Hypermnesia for pictures and words: The role of recall modality (pictures vs. words). *Bulletin of the Psychonomic Society, 16,* 258–260.

Erdelyi, M. (1984). The recovery of unconscious (inaccessible) memories: Laboratory studies of hypermnesia. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 18, pp. 95–127). New York: Academic Press.

Erdelyi, M. H., & Becker, J. (1974). Hypermnesia for pictures: Incremental memory for pictures but not for 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 hypermnesia: Is a rose a rose a rose? *Cognition, 4,* 311–319.

Erdelyi, M. H., & Kleinbard, J. (1978). Has Ebbinghaus decayed with time?: The growth of recall (hypermnesia) over days. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 275–289.

Erdelyi, M. H., & Stein, J. B. (1981). Recognition hypermnesia: The growth of recognition memory (d') over time with repeated testing. *Cognition, 9,* 23–33.

Hochhaus, L. (1972). A table for the calculation of d' and β. *Psychological Bulletin, 77,* 375–376.

Hogan, R. M., & Kintsch, W. (1971). Differential effects of study and test trials on long-term recognition and recall. *Journal of Verbal Learning and Verbal Behavior, 10,* 562–567.

Indow, T., & Togano, K. (1970). On retrieving sequence from long term memory. *Psychological Review, 77,* 317–331.

Kintsch, W. (1970). Models for free recall and recognition. In D. A. Norman (Ed.), *Models of human memory* (pp. 331–373). New York: Academic Press.

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

Madigan, S. (1976). Reminiscence and item recovery in free recall. *Memory & Cognition, 4,* 233–236.

Madigan, S., & Lawrence, V. (1980). Factors affecting item recovery and hypermnesia in free recall. *American Journal of Psychology, 93,* 489–504.

McDaniel, M. D., & Masson, M. E. J. (1985). Altering memory representations through retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 371–385.

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monographs, 76*(1, Pt. 2).

Paris, S. G. (1978). Memory organization during children's repeated recall. *Developmental Psychology, 14,* 99–106.

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

Payne, D. G. (1987). Hypermnesia and reminiscence in recall: A historical and empirical review. *Psychological Bulletin, 101,* 5–27.

Popkin, S. J., & Small, M. Y. (1979). Hypermnesia and the role of imagery. *Bulletin of the Psychonomic Society, 13,* 378–380.

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: Advances in research and theory* (Vol. 14, pp. 207–262). New York: Academic Press.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88,* 93–134.

Roediger, H. L. (1982). Hypermnesia: The importance of recall time and asymptotic level of recall. *Journal of Verbal Learning and Verbal Behavior, 21,* 662–665.

Roediger, H. L., & Payne, D. G. (1982). Hypermnesia: The role of repeated testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8,* 66–72.

Roediger, H. L., & Payne, D. G. (1985). Recall criterion does not affect recall level or hypermnesia: A puzzle for generate/recognize theories. *Memory & Cognition, 13,* 1–7.

Roediger, H. L., Payne, D. G., Gillespie, G. L., & Lean, D. S. (1982). Hypermnesia as determined by level of recall. *Journal of Verbal Learning and Verbal Behavior, 21,* 635–655.

Roediger, H. L., & Thorpe, L. A. (1978). The role of recall time in producing hypermnesia. *Memory & Cognition, 6,* 296–305.

Rosner, S. R. (1970). The effects of presentation and recall trials on organization in multitrial free recall. *Journal of Verbal Learning and Verbal Behavior, 9,* 69–74.

Schacter, D. L. (1983). Feeling of knowing in episodic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9,* 39–54.

Shapiro, S. R., & Erdelyi, M. H. (1974). Hypermnesia for pictures but not for words. *Journal of Experimental Psychology, 103,* 1218–1219.

Shiffrin, R. M. (1970). Memory search. In D. A. Norman (Ed.), *Models of human memory* (pp. 375–447). New York: Academic Press.

Standing, L., Conezio, J., & Haber, R. N. (1970). Perception and memory for pictures: Single-trial learning of 2560 visual stimuli. *Psychonomic Science, 19,* 73–74.

Thompson, C. P., Wenger, S. K., & Bartling, C. A. (1978). How recall facilitates subsequent recall: A reappraisal. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 210–221.

Tulving, E. (1967). The effects of presentation and recall in free recall learning. *Journal of Verbal Learning and Verbal Behavior, 6,* 175–184.

Tulving, E. (1968). Theoretical issues in free recall. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory* (pp. 2–36). Englewood Cliffs, NJ: Prentice-Hall.

Tulving, E. (1974). Cue-dependent forgetting. *American Scientist, 62*, 74–82.

Wenger, S. K., Thompson, C. P., & Bartling, C. A. (1980). Recall facilitates subsequent recognition. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 135–144.

Yarmey, A. D. (1976). Hypermnesia for pictures but not for concrete or abstract words. *Bulletin of the Psychonomic Society, 8*, 115–117.