Recall criterion does not affect recall level or hypermnesia: A puzzle for generate/recognize theories

HENRY L. ROEDIGER, III, and DAVID G. PAYNE
Purdue University, West Lafayette, Indiana

Subjects studied a mixed list of 70 words that varied in imagery value and then received three successive tests. Also varied were instructions given to subjects prior to list presentation (imagery or semantic rehearsal) and the type of recall test (standard free recall, an uninhibited-recall procedure in which subjects were encouraged to free associate and to guess while recalling the list, and a forced-recall condition in which they were also told to write a large number of responses to fill the allotted spaces). Recall improved across the three tests in all conditions, but the improvement was greater for high- than for low-imagery words. In addition, hypermnesia (the improved recall across tests) was shown to occur following semantic rehearsal instructions as well as imagery instructions and to occur with low-imagery words, contrary to the imagery hypothesis of the effect. Most importantly, the large variation in recall criterion produced by manipulating instructions at test (as measured by intrusions) did not affect the overall level of correct recall or the magnitude of improvements across tests. Apparently, the assumption of generate-recognize theories that people generate much more information in free recall than they produce (due to a stringent criterion for recognition of the generated material) is false.

Hypermnesia refers to the improvement in recall across repeated tests. The term was introduced into the modern literature by Erdelyi and Becker (1974) as a replacement for the term reminiscence (used by Ballard, 1913, in describing his experiments), because reminiscence had come to have multiple meanings. Some researchers used reminiscence to refer to absolute improvements across tests, whereas others applied it to recall of events that were forgotten on one test but recovered on a later one. (This latter sense was the one originally used by Ballard.) In this paper we retain the usage, suggested by Roediger and Thorpe (1978), of reminiscence as recovery of previously forgotten items on a later test (intertest recovery) and hypermnesia as the absolute increase of recall across tests (when intertest recovery exceeds intertest forgetting).

The finding of improvements in recall across repeated tests has suffered several reversals of fortune in this century, with some researchers finding it and others not (see Woodworth, 1929, pp. 63-68, and McGeoch & Irion, 1952, pp. 140-148, for the early history). However, in recent years the procedure introduced by Erdelyi and his colleagues has provided a reliable method for demonstrating and studying the phenomenon (e.g., Erdelyi & Becker, 1974; Erdelyi, Buschke, & Finkelstein, 1977; Erdelyi, Finkelstein, Herrell, Miller, & Thomas, 1976). Typically, subjects study pictures or words and then receive three

successive recall tests of 7-min duration each. Some of the main findings to emerge from Erdelyi's research cited above are that: (1) recall improves for pictures but not for concrete words across the three tests; (2) study of concrete words following instructions to image the words produces hypermnesia; and (3) generating words at study from a question (e.g., "This longish wooden object is used by baseball players to hit the ball. What is it?") also produces the effect. Erdelyi hypothesized that imaginal coding is the critical factor producing hypermnesia and that the intermittent success of researchers in obtaining this effect is due to inconsistent use of materials that foster imaginal coding. (To account for the third finding above, one must assume that generation of concepts involves imaging them.)

Hypermnesia has received theoretical treatments aside from the imagery hypothesis offered by Erdelyi and his co-workers. For example, Roediger and Payne (1982), Roediger and Thorpe (1978), and Roediger, Payne, Gillespie, and Lean (1982) have argued that considerations of recall practice, recall time, and level of recall within the hypermnesia experiment are critical in determining whether or not the effect will be obtained. The general argument is that hypermnesia is directly related to level of recall irrespective of the coding format of the materials, although this assertion must be qualified (see Roediger, 1982). The present experiment was designed in part to provide evidence bearing on the imagery and recall level hypotheses. A second purpose was to clear up several empirical discrepancies that have appeared in the literature on hypermnesia. Each will be reviewed below, but briefly the three issues are: (1) whether or not

hypermnesia can be found with concrete words when subjects are not given instructions to form images; (2) whether or not semantic elaboration instructions will also lead to hypermnesia; and (3) whether or not hypermnesia occurs with low-imagery words. A fourth issue is of more general interest to theories of recall, because it permits a test of the common assumption that recall involves a recognition stage employing a high criterion, so that people must be relatively certain that a memory is accurate before they will make it public. We ask whether or not one can show criterion effects in correct recall ("hits") by having subjects vary their response criteria to produce few or many intrusions ("false alarms"). Our experiment provides unambiguous answers to all four of these questions. We review next the background for the four issues motivating this research, discussing each in turn.

First, does the type of recall test affect whether or not hypermnesia occurs for high-imagery (concrete) words when subjects are not told to image the referents of the words at study? One consistent difference between research in our laboratory and that of Erdelyi and his colleagues has been in finding hypermnesia for high-imagery words; we have (e.g., Roediger & Thorpe, 1978) and they have not (e.g., Erdelyi & Becker, 1974). One possible reason is that in Erdelyi's studies a forced-recall procedure has been employed in which subjects are required to recall a constant (and large) number of words on each test, guessing when necessary. This procedure controls for criterion shifts across the three tests, because subjects must produce the same number of responses on each test. The interest in our experiments was in measuring the number of items recalled per unit time, which we thought would be hampered if subjects were constrained to guess large numbers of items on each test. Thus, we used a standard free-recall procedure and required subjects to draw lines under the last item written at specified times during each test. We satisfied ourselves that criterion shifts were not a problem by noting that intrusions were relatively rare and did not increase across tests. Nonetheless, it is possible that our finding of hypermnesia for concrete words (when subjects were not given imagery instructions) depends on criterion shifts in free recall and could not be obtained with the forced-recall procedure. To determine this, in the present experiment we employed variations in the type of recall test following presentation of words to determine whether this factor is important in producing hypermnesia for concrete words.

A second issue addressed in the current experiment was the effect of study instructions on producing hypermnesia. As already mentioned, Erdelyi et al. (1976) showed hypermnesia for concrete words when subjects were given imagery instructions prior to studying the words. However, Roediger and Thorpe (1978) found hypermnesia for both high- and low-imagery words when subjects were given semantic elaboration instructions. (In this condition, they were told to think of the meaning of the words by thinking of associates or making up a story.) Belmore

(1981) also found hypermnesia for words studied under semantic elaboration instructions, but her results showed small gains across tests and were somewhat variable from experiment to experiment. Since she used forced recall and Roediger and Thorpe (1978) used free recall, we decided to manipulate type of study instruction (imagery or semantic elaboration) orthogonally to the type of test (free or forced recall) to see whether type of test caused these variable results.

The issue of whether or not concrete words produce hypermnesia may not really be critical evidence with regard to the imagery hypothesis, despite the fact that failing to find hypermnesia with high-imagery words (but finding it with their pictorial representations) is what led to the hypothesis. The reason is that one can reasonably suppose that processing high-imagery words involves some form of imaginal coding, as Paivio (1969) proposed in his dual coding hypothesis. A question of more importance for the imagery hypothesis is whether or not hypermnesia occurs for low-imagery words. Of course, subjects can form images even to low-imagery words, but the presumption is that they do not take the considerable time and effort to do so unless specifically instructed to (Paivio, 1969).

Roediger and Thorpe (1978, Experiment 2) did indeed show hypermnesia for low-imagery words presented slowly (5 sec/item) when subjects were given imagery instructions prior to presentation. However, they also showed a similar pattern of recall for low-imagery words when subjects were given semantic elaboration instructions. This last finding seems incompatible with the idea that imaginal coding is necessary for hypermnesia, but the implication of the finding is cloudy, because others have failed to find hypermnesia with low-imagery words. Although Hasher, Riebman, and Wren (1976) also reported reliable improvements between two tests for low-imagery words, Yarmey (1976) did not. (In fact, Yarmey found no hypermnesia for high-imagery words in his forced-recall tests.) Popkin and Small (1979) noted that these discrepancies might be accounted for by the fact that the actual imagery values of the supposedly "low-imagery" words were rather different in these studies. Roediger and Thorpe (1978, Experiment 2) equated low- and high-imagery words on meaningfulness, word length, and frequency of usage, which constrained the values of their low-imagery words. Their low-imagery words averaged 4.3 on Paivio, Yuille, and Madigan's (1968) 7-point scale, on which 7 represents the highest value. Hasher et al. (1976) used low-imagery words with a mean rating of 4.1. (Yarmey's low-imagery words were simply said to differ significantly in imagery and concreteness from very high-imagery words.) Popkin and Small used low-imagery words that varied between ratings of 1 and 3 in the Paivio et al. norms, and they failed to find any hypermnesia. They concluded that hypermnesia is not observed with truly low-imagery words.[1] A third purpose of these experiments was to obtain further evidence on this point by varying the imagery value of the words.

A fourth aim of this research was to test a proposition inherent in many theories of recall, but that has received little empirical attention. Starting at least with Deese (1959), many theorists have assumed that recall is a sequential process involving at least two stages, one by which information is generated from memory and a second for deciding whether generated information should be overtly produced (e.g., Anderson & Bower, 1972; Kintsch, 1970; Peterson, 1967; Shiffrin, 1970). In such generate/recognize theories, the recognition process is usually regarded as stringent, because subjects typically make few intrusions in recall (at least of arbitrary lists of words and pictures). If this is so, then the decision process must have a fairly high threshold for accepting a response, which leaves open the possibility that subjects generate many responses that are not produced. This leads to the prediction that if subjects were induced to relax their recall criteria, performance would be improved (but at the expense of more intrusions).

Whether or not one specifically endorses generate/recognize theories, the implicit assumption that recall involves a criterion for acceptability of candidate responses is ingrained in our thinking. Consider the problem of whether or not hypnosis improves memory retrieval (Smith, 1983). Some students do indeed show that subjects who are hypnotized recall more under this state than do control subjects given similar instructions but not hypnotized (e.g., Dywan & Bowers, 1983). However, in this study, subjects in the hypnosis condition also produced more false memories, leading the researchers to conclude that the higher rate of recall under hypnosis was probably an artifact produced by a lowering of the criterion for acceptability under hypnosis. The underlying assumption is that if the high criterion for recall is relaxed, a greater number of hits will be produced, but only at the expense of a concomitant increase in false alarms. Klatzky and Erdelyi (in press) explicated this problem well, within the framework of signal detection theory.

This logic is perfectly reasonable, but has never been tested extensively. However, several studies have provided some reason to doubt it. Cofer (1967) presented subjects with 15-item lists and, after a period of free recall, employed a further period of forced recall in which subjects were required to produce enough responses to match the list length. Despite the production of many responses during this period, subjects rarely recalled correct items, leading Cofer to conclude that "free recall approximately exhausts the list items that are in storage" (p. 201). Similarly, Keppel and Mallory (1969) gave one group of subjects instructions to guess in a free-recall situation and compared their performance with that of two other groups. One of these was given standard free-recall instructions that did not mention guessing, whereas the other was given similar instructions but was also warned against guessing. Subjects given instructions to guess did indeed provide more intrusions than the other two groups (which did not differ on this measure), but produced no more correctly recalled items. However, the differences in intru-

sion rates (indexing a criterion shift) were rather small. In a similar study, Bousfield and Rosner (1970) compared subjects' performance under standard free-recall instructions with that of others under an "uninhibited"-recall condition in which subjects were told to say any words that occurred to them in the process of attempting recall, even though they might make errors. Subjects were given five study/test trials. Bousfield and Rosner presented data for only Trials 1 and 5. There was no difference between groups in correct recall on Trial 1, despite a reliable increase in intrusions for the uninhibited-recall group. They did report a significant benefit between groups in the number of items recalled on Trial 5, but it was fairly small.

The experiment reported here provided another opportunity to evaluate the question of whether or not a high criterion normally exists in recall and operates to dampen the number of memories reported. We varied recall criteria over a much larger range than in previous studies, as indexed by the number of intrusions. Subjects were shown a 70-word list and then given three successive tests in one of three recall conditions: standard free recall, Erdelyi and Becker's (1974) forced-recall condition, and Bousfield and Rosner's (1970) uninhibited-recall condition. In addition, half the 70 words were high-imagery words, and half were low-imagery words. Finally, half the subjects were given instructions prior to presentation to use imaginal coding, whereas the other subjects were instructed to use semantic rehearsal. Thus, to summarize, the experiment was designed to determine: (1) whether type of recall test (free or forced) affects hypermnesia for words; (2) whether hypermnesia could be obtained for low-imagery words; (3) whether hypermnesia could be obtained with semantic elaboration instructions; and (4) the more general issue of whether relaxing recall criteria produces an increase in correctly recalled items.

METHOD

Subjects
The subjects were 120 Purdue University undergraduates who served as one method of fulfilling an introductory psychology course requirement.

Design
Two between-subject factors and two within-subject factors were included in the design. Varied between subjects were the two types of instruction for how to learn the words (imagery and semantic rehearsal) and three types of recall tests (free recall, forced recall, and uninhibited recall). Twenty subjects participated in each of these six between-subject conditions. The within-subject variables were the type of material (high- and low-imagery words) and the number of the test (first, second, and third). Thus, the entire design can be considered a 2 × 3 × 2 × 3 mixed design.

Materials
Two sets of 35 words were selected from the norms of Paivio et al. (1968) to differ widely on the imagery and concreteness scales, but to be of equal frequency of occurrence in the language. The high-imagery words had mean values of 6.63 and 6.80 on the imagery and concreteness scales (which vary from 1-7), whereas the comparable figures for the low-imagery words were 3.24 and 2.63. (These values were substantially lower than those for items used in Roediger & Thorpe, 1978, Experiment 2.) All words were A

or AA in the Thorndike and Lorge (1944) word counts. High-imagery words were somewhat higher than low-imagery words in meaningfulness (mean values of 7.05 to 5.73) and contained, on average, about one fewer letter (5.1 letters to 6.4).

Procedure

The subjects were tested in small groups, which were randomly assigned to the six between-subject conditions formed by variation in study instructions and the type of recall test. At the beginning of the session, the subjects were told that they would be shown a relatively long list of words to study for a later test. They were also told that they would be provided a specific strategy to enhance recall of the list. The subjects were then given either the imaginal encoding or semantic rehearsal instructions. In the former, they were told to try to form "a clear and vivid image" of the referent of each word as it was presented, although they were informed that this might be difficult in some cases. In the other condition, the subjects were told to repeat the word to themselves until the next word came up, but also to try to think of the meaning of the word and associates of the word as they did so, rather than simply using rote repetition. Examples of these coding operations were given using nonlist words. The subjects were told nothing of the type or number of tests during the initial instructions. After the instructions were read and questions (if any) answered, the subjects were shown the lists of words via a Kodak Carousel slide projector at a 5-sec rate.

Following presentation of the list, the subjects were given instructions corresponding to the type of recall test they were to be given. Those given standard free-recall instructions were told to remember as many words as they could in any order, but were warned not to guess wildly. The subjects given uninhibited-recall instructions (modeled after those of Bousfield & Rosner, 1970) were told to recall as many list items as possible in any order, but in addition were told to write down whatever other words came to mind while they were attempting recall. They were told to free associate while recalling the list and to produce many responses, regardless of the nature of those responses and even if they knew they were only guessing. In the forced-recall condition, the subjects were given similar instructions, but told that they had to write down 50 items on their test sheets, even if they had to guess items to fill the sheet. They were told that they would be warned as the end of the test neared, so that they could complete the 50 spaces on their sheets.

The subjects in all three groups were informed that they would be instructed to draw a line after the last item that they had written during each minute of the recall test, to permit measurement of the rate of recall. (They were also told to draw a line even if no additional words had been recalled since the previous instruction.) All

groups were given three 8-min tests, with warnings given at 2 and 1 min of the remaining time. (The forced-recall group was given an additional warning with 3 min remaining.) The subjects were told that they would have 8 min for the test and that they should keep trying during the entire test, even when it seemed difficult to recall more items. The first recall test began 2 min after the list had been presented. At the end of the first test, recall sheets were collected and the subjects were given a short break, during which they received additional instructions. These instructions informed the subjects that they were to be given another test like the first, and the test instructions for their condition were repeated in shortened form. The subjects were also told that they should try to improve total recall on the second test by recalling the words that had already been recalled on the first test and as many new words as possible. Following the instructions, the subjects were allowed to relax for the remainder of the 2-min intertest interval before the second test commenced.

At the end of the second 8-min test, the subjects were told that they would have one more rest period and then be asked to recall the items a third and final time. The instructions were similar to the ones given prior to the second test. After the third test, the subjects were debriefed as to the purpose of the experiment.

RESULTS AND DISCUSSION

The factor of type of study instructions (imagery vs. semantic rehearsal) did not produce a significant main effect ($F < 1$, MSe = 84.51) and did not participate in any significant interactions. Therefore, to simplify matters, the results are combined across this factor. Mean recall on each of the three tests as a function of the other variables is presented in Table 1, along with the difference between recall on the third and first tests ($T_3 - T_1$, an index of hypermnesia) and the percentage of improvement relative to recall on the first test. Each mean in the first nine rows is based on 40 observations; the means in the bottom three rows are based on all 120 subjects.

Examination of the table reveals consistent improvement across the three tests, indicating that hypermnesia was obtained [$F(2,228) = 68.4$, MSe = 2.24]. (All effects deemed statistically significant reach at least the .05 level of confidence.) The effect of imagery value of the words was also significant [$F(1,114) = 562.7$, MSe =

Table 1
Mean Number of Items Recalled as a Function of Type of Material, Type of Test, and Number of Tests

Type of Test	Imagery Value	Tests			Hypermnesia	
		1	2	3	$T_3 - T_1$	Percent Improvement
Free Recall	High	16.42	17.65	18.50	2.08	12.7
	Low	8.68	8.92	9.70	1.02	11.8
	Total	25.10	26.57	28.20	3.10	12.4
Uninhibited Recall	High	15.52	17.32	18.30	2.78	17.9
	Low	8.75	9.05	10.10	1.35	15.4
	Total	24.27	26.37	28.40	4.13	17.0
Forced Recall	High	16.95	17.18	18.38	1.43	8.4
	Low	8.12	8.80	9.05	.93	11.4
	Total	25.07	25.98	27.43	2.36	9.4
All Tests Combined	High	16.30	17.38	18.39	2.09	12.8
	Low	8.52	8.92	9.62	1.10	12.9
	Total	24.82	26.30	28.01	3.19	12.9

Note—"$T_3 - T_1$" = Test 3 − Test 1.

22.4], with high-imagery words being recalled better than low-imagery words (17.36 to 9.02). However, the main effect of type of recall test (free, uninhibited, forced) did not produce a reliable effect on recall [$F(2,114) < 1$, $MSe = 84.51$]. However, the number of tests × materials interaction did prove reliable [$F(2,228) = 7.15$, $MSe = 2.15$], with greater hypermnesia occurring for high-imagery words than for low-imagery words, as shown in the combined data at the bottom of Table 1. No other interactions were significant.

The data permit clear answers to the four questions motivating the research. First, does the type of recall test (particularly free vs. forced) affect hypermnesia for words? The answer is no. Hypermnesia of roughly the same magnitude was found on all three types of test, indicating that this methodological difference in prior research is not responsible for variable findings that are occasionally obtained. Second, does hypermnesia occur for words studied under semantic rehearsal instructions with both free- and forced-recall tests? The answer here is yes, because type of study instruction produced no reliable main effect and participated in no significant interactions. Combining across imagery value of words and type of test, the subjects given imagery study instructions recalled 24.70, 26.32, and 27.78 words across the three tests, whereas those having semantic rehearsal instructions recalled 24.93, 26.30, and 28.23. Type of instruction did not interact with the imagery value of the words, either. Thus, contrary to the imagery hypothesis, semantic elaboration instructions produced hypermnesia both for low- and high-imagery words.

The third question, does hypermnesia occur for low-imagery words, was also answered affirmatively. Across the three tests, recall improved even when low-imagery words were analyzed separately [$F(2,228) = 17.24$, $MSe = 1.63$]. High-imagery words produced greater absolute hypermnesia than did low-imagery words, with the advantage on the third test over the first being 2.09 for high-imagery words and 1.10 for low-imagery words. However, when the percentage of gain across tests is calculated by dividing the absolute gain by mean recall on the first test, the figures are shown to be almost identical (12.8% and 12.9%).

The fourth issue raised is whether or not a criterion problem exists under standard free-recall conditions that operates to depress recall. The three tests (free, forced, and uninhibited) were supposed to vary recall criteria as

measured by the number of intrusions (which they did), but there was no resulting effect on recall "hits." Relevant data are shown in Table 2, in which mean recall on each of the three tests is presented (combining across the other variables) for each recall instruction condition. Also presented are mean recall across tests, mean number of intrusions, and the amount of hypermnesia as indexed by the difference in recall between the third and first tests. The mean number of intrusions varied from 2.47 in free recall to 23.84 in forced recall, but despite this almost tenfold increase in intrusions, mean correct recall in the three conditions averaged across tests differed by less than half an item (and that even favored free recall). If one examines the total number of unique items recalled over the 24-min period, the outcome remains the same. The subjects recalled 31.15, 30.95, and 30.90 items in the free-, uninhibited-, and forced-recall conditions, respectively.

The amount of hypermnesia obtained is also unrelated to recall criterion. The improvement across tests was slightly greater with uninhibited recall than with free recall, but was lowest with forced recall (for which the most intrusions occurred). (As mentioned, the F for type of test was less than unity.) The inescapable conclusion is that neither level of recall nor hypermnesia depends on recall criteria. Admittedly, this conclusion is based on a failure to reject the null hypothesis, but (1) a fairly large number of observations are involved in designs that showed reliable effects of several other variables, (2) shifts in recall criteria, indexed by the number of intrusions varied over a wide range, and (3) the data did not even provide a trend in the predicted direction. Too, other reports have yielded similar results (e.g., Keppel & Mallory, 1969). Contrary to the assumption of generate/recognize theories, the implicit assumption of other theories, and common sense, people do not seem to generate more information than they can recall in a standard free-recall situation.

IMPLICATIONS

Besides helping to resolve some empirical issues, the present results have additional theoretical implications. Certain aspects of the results disagree with predictions from Erdelyi and Becker's (1974) proposal that imaginal coding is the critical factor in producing hypermnesia. First, hypermnesia was obtained for low-imagery words;

Table 2
Mean Recall, Intrusions, and Hypermnesia for the Three Types of Test

| Type of Test | Test | | | Mean | Intrusions | $T_3 - T_1$ |
	1	2	3			
Free Recall	25.10	26.58	28.20	26.63	2.47	3.10
Uninhibited Recall	24.28	26.35	28.40	26.35	9.60	4.12
Forced Recall	25.08	25.98	27.42	26.16	23.84	2.34
Mean	24.82	26.30	28.01	26.38		3.19

Note—"$T_3 - T_1$" = same as in Table 1.

second, it was obtained when subjects were instructed to remember words by being given semantic rehearsal instructions; and third, the combination of these factors (low-imagery words with semantic rehearsal instructions) still produced the effect. These results and conclusions reinforce those obtained from similar experiments (Belmore, 1981; Roediger et al., 1982; Roediger & Thorpe, 1978). They are also in general agreement with Roediger et al.'s hypothesis that recall level determines hypermnesia, because in this experiment hypermnesia was affected by variables that affected recall level (the imagery value of the words), but not by variables that did not affect recall level (the type of recall test, or the type of study instructions).

Although our data are generally inconsistent with the imagery hypothesis, they are probably not strong reasons to refute it. First, most are based on failures to reject the null hypothesis. Second, although we find no difference between semantic elaboration instructions and imaginal coding instructions in producing hypermnesia, we also find no evidence that the subjects given these different instructions varied in any way. Perhaps the manipulation simply failed. Third, the mixing of high- and low-imagery words together in the list may have encouraged the subjects to use imaginal coding. We tend to doubt these objections, but cannot rule them out on the basis of current evidence. However, other evidence produced by Payne (1984) shows that imaginal coding does play a role in producing hypermnesia beyond simply raising level of recall. In his experiments, recall of pictures and concrete words was compared at different overall levels by varying presentation time and number of presentations for the two types of material. Under some conditions, he found greater hypermnesia for pictures than for words, even when overall recall of words was higher than that of pictures. This evidence clearly shows a role for imaginal coding over and above that due to enhanced recall in producing hypermnesia.

The wider implication of the present results, to state the proposition in its most extreme form, is that there is no criterion effect in recall. In the forced-recall condition, the intrusion rate was varied by almost a factor of 10 relative to that in the free-recall condition (from 2.47 to 23.84), and yet the hit rates (the number of items correctly recalled) did not vary at all. Contrary to at least the spirit of generate/recognize theories, subjects apparently are not normally generating correct candidate responses in free recall that they reject on an implicit recognition test as being incorrect, since encouraging them to spew these out produces no more correct responses than are found in standard free recall.

Of course, it might be possible to show such effects if subjects in the standard comparison condition were in some way discouraged from producing maximum performance. But it seems remarkable that, relative to standard free recall with instructions discouraging guessing, telling subjects to guess wildly (the uninhibited-recall condition) or even forcing them to write down responses (the forced-recall condition) produces no gain in performance. However surprising, this result agrees with the findings of Keppel and Mallory (1969) with free recall, as well as with the outcomes obtained by Scheible and Underwood (1954) with serial learning and Underwood and Ekstrand (1967) with paired-associate learning. In a review of related work, Weiner (1966a) found no apparent effect of varying motivation on retrieval. That is, in between-subject designs in which subjects are offered various rewards for performing well, their performance does not differ from that of control subjects simply asked to recall as well as possible (see Weiner, 1966b, Experiments VI-IX). In an unpublished experiment performed by the first author, subjects were presented with a list of unrelated words and then were given free-recall instructions with the difference between conditions (manipulated between subjects) being in the monetary reward for recalling the words: 0, 5, or 50 cents per word recalled. No differences were observed in the number of items recalled.

One might question whether these findings concerning response criteria will generalize to types of material other than word lists. Work reported by Britton, Meyer, Hodge, and Glynn (1980) permits an affirmative answer, at least tentatively. In Experiment 4 of their report, subjects were presented with prose passages that varied in content structure. At test, subjects were given free-recall instructions that imposed either a low criterion (write down everything and guess) or a high criterion (write down only important material). The criterion manipulation reliably affected the number of words produced, but had no significant effect on target recall, although the low-criterion group did slightly better. However, both groups performed no better (in fact, slightly worse) than subjects recalling the same passages under standard free-recall conditions.

Findings such as the ones reported and reviewed here should caution against the facile interpretation, which is all but automatic, of arguing that differences in recall between conditions should be discounted if intrusion rates also vary in the same direction. As mentioned in the introduction, the effect of hypnosis on memory retrieval is one area in which this inference is routinely made (e.g., Klatzky & Erdelyi, in press). It may well be that the occasional gains under hypnosis are real ones and are not due merely to criterion shifts, since (to our knowledge) no one has shown that criterion shifts per se (relative to standard free-recall conditions) can improve retrieval. What is needed is an experiment in which recall criteria are varied across groups of subjects who recall while under hypnosis. Perhaps changes in recall criteria do affect hypnotic recall, as suggested by Klatzky and Erdelyi and others (Dywan & Bowers, 1983).

In an interesting experiment that tested the effects of hypnosis on retrieval when people were interviewed by police officers trained in the technique, Geiselman, Fisher, MacKinnon, and Holland (in press) reported increases from hypnosis relative to a control condition, with no concomitant increase in the number of false memories pro-

duced. (Intrusions were very low in all conditions.) However, Geiselman et al. suggested that the policemen's interviewing strategy provided effective retrieval cues and that these, rather than hypnosis, probably enhanced recall. Other interpretive problems also exist in evaluating the effects of hypnosis (see Smith, 1983), even if, in light of further research, criterion shifts turn out not to be one of them. It is rather puzzling that the problem of recall criteria has received so little attention, since some of our most popular theories make strong assumptions about their role, as do commonsense folk theories. Apparently, these popular and reasonable assumptions of a strict criterion operating to depress recall are false.

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(No. 2).

BELMORE, S. M. (1981). The role of imagery and semantic elaboration in hypermnesia for words. *Journal of Experimental Psychology: Human Learning and Memory*, **7**, 204-215.

BOUSFIELD, W. A., & ROSNER, S. R. (1970). Free versus uninhibited recall. *Psychonomic Science*, **20**, 75-76.

BRITTON, B. K., MEYER, B. J., HODGE, M. H., & GLYNN, S. M. (1980). Effects of organization of text on memory: Tests of retrieval and response criterion hypotheses. *Journal of Experimental Psychology: Human Learning and Memory*, **6**, 620-629.

COFER, C. N. (1967). Does conceptual organization influence the amount retained in immediate free recall? In B. Kleinmuntz (Ed.), *Concepts and the structure of memory*. New York: Wiley.

DEESE, J. (1959). Influence of inter-item associative strength upon immediate free recall. *Psychological Reports*, **5**, 305-312.

DYWAN, J., & BOWERS, K. (1983). The use of hypnosis to enhance recall. *Science*, **222**, 184-185.

ERDELYI, M. H., & BECKER, J. (1974). Hypermnesia for pictures: Incremental memory for pictures but not words in multiple recall trials. *Cognitive Psychology*, **6**, 159-171.

ERDELYI, M. H., BUSCHKE, H., & FINKELSTEIN, S. (1977). Hypermnesia for Socratic stimuli: The growth of recall for an internally generated memory list abstracted from a series of riddles. *Memory & Cognition*, **5**, 283-286.

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.

GEISELMAN, R. E., FISHER, R. P., MACKINNON, D. P., & HOLLAND, H. L. (in press). Eyewitness memory enhancement in the police interview: Cognitive retrieval mnemonics versus hypnosis. *Journal of Applied Psychology*.

HASHER, L., RIEBMAN, B., & WREN, F. (1976). Imagery and the retention of free recall learning. *Journal of Experimental Psychology: Human Learning and Memory*, **2**, 172-181.

KEPPEL, G., & MALLORY, W. A. (1969). Presentation rate and instructions to guess in free recall. *Journal of Experimental Psychology*, **79**, 269-275.

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.

KLATZKY, R. L., & ERDELYI, M. (in press). The response criterion problem in tests of hypnosis and memory. *International Journal of Clinical and Experimental Hypnosis*.

MCGEOCH, J. A., & IRION, A. L. (1952). *The psychology of human learning*. New York: Longmans, Green.

PAIVIO, A. (1969). Mental imagery in associative learning and memory. *Psychological Review*, **76**, 241-263.

PAIVIO, A., YUILLE, J. C., & MADIGAN, S. (1968). Concreteness, imagery, and meaningfulness values for 925 concrete nouns. *Journal of Experimental Psychology Monograph Supplement*, **76**(1, Pt. 2).

PAYNE, D. G. (1984). *An investigation of the imaginal coding and recall level interpretations of the hypermnesia phenomenon*. Unpublished doctoral dissertation, Purdue University, West Lafayette, IN.

PETERSON, L. R. (1967). Search and judgment in memory. In B. Kleinmuntz (Ed.), *Concepts and the structure of memory*. New York: Wiley.

POPKIN, S. J., & SMALL, M. Y. (1979). Hypermnesia and the role of imagery. *Bulletin of the Psychonomic Society*, **13**, 378-380.

ROEDIGER, H. L. (1982). Rejoinder. 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. P. (1982). Hypermnesia: The role of repeated testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **8**, 66-72.

ROEDIGER, H. L., PAYNE, D. G., GILLESPIE, G. L., & LEAN, D. (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.

SCHEIBLE, H., & UNDERWOOD, B. J. (1954). The role of overt errors in serial learning. *Journal of Experimental Psychology*, **47**, 160-162.

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

SMITH, M. C. (1983). Hypnotic memory enhancement of witnesses: Does it work? *Psychological Bulletin*, **94**, 387-407.

THORNDIKE, E. L., & LORGE, I. (1944). *The teacher's wordbook of 30,000 words*. New York: Teachers College Press, Columbia University.

UNDERWOOD, B. J., & EKSTRAND, B. R. (1967). Effect of distributed practice on paired-associate learning. *Journal of Experimental Psychology Monographs*, **73**(3, Pt. 2).

WEINER, B. (1966a). Effects of motivation on the availability and retrieval of memory traces. *Psychological Bulletin*, **65**, 24-37.

WEINER, B. (1966b). Motivation and memory. *Psychological Monographs: General and Applied*, **80**(18, No. 626).

WOODWORTH, R. S. (1929). *Psychology*. New York: Holt.

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

NOTE

1. Of course, if the high- and low-imagery words were not equated on other variables (frequency of occurrence, length, meaningfulness), then it is logically possible that one of these features was confounded with imagery value and may be responsible for the failure to obtain hypermnesia.