Current Issues
in Cognitive Processes

The Tulane Flowerree
Symposium on Cognition

Edited by
Chizuko Izawa
Tulane University
Hypermnesia: Improvements in Recall With Repeated Testing

Henry L. Roediger, III  
*Rice University*

Bradford H. Challis  
*Purdue University*

Memory is never so perplexing a capacity as when it fails. Everyone has had the experience of trying repeatedly to recall something well known, but failing utterly. Just as curious is the experience of the forgotten fact or name suddenly reappearing at some later point, even without a conscious intention to remember it. In his monumental experimental monograph, *Über Das Gedächtnis*, Ebbinghaus (1885/1964) noted that “Names, faces, bits of knowledge and experience that had seemed lost for years suddenly appear before the mind, especially in dreams, with every detail present and in great vividness; and it is hard to see whence they came and how they managed to keep hidden so well in the meantime” (p. 62).

The phenomenon of spontaneous remembering after retrieval failures is difficult to study empirically. One method sometimes used to investigate such memory failures and recoveries involves having people keep diaries of such experiences (Reason & Lucas, 1984). This method provides descriptive information and may be useful in the first stages of investigating the phenomenon. However, most researchers have preferred experimental methods, dating back to the turn of the century. The typical strategy has been to provide subjects with material to remember and then to test them repeatedly with no intervening study. The usual finding is that people remember material on later tests that they failed to produce on earlier tests. Because the nature of the external retrieval conditions typically does not change between tests, these represent cases of spontaneous remembering. Ballard (1913) provided the first well known series of experiments demonstrating these phenomena. Following presentation of different types of material to school children, he gave them repeated tests at various intervals of time since learning (but with no intervening study opportunities). He discovered that the children almost always recalled
material on later tests that they could not recall on earlier tests, and that frequently the total amount recalled improved on the later tests. (That is, recovery of material between tests outweighed forgetting.)

REMINISCENCE AND HYPERMNESIA

We describe later experiments by W. Brown (1923) to illustrate the basic phenomena from this sort of experiment and to introduce some terminology. Brown asked the basic question that formed the title of his paper, “To what extent is memory measured by a single recall?” Brown performed two experiments. In the first, he asked college students to recall as many of the 48 states as possible during a 5-minute period. A half hour later, without any warning, he gave them the same task again. The interest was in determining differences in performance between the two tests. The second experiment was similar, except that in this case students were given 48 words to memorize, with each word presented four times. Once again, an initial 5-minute recall occurred immediately after learning, and second test followed after 30 minutes.

The results from Brown’s experiments are presented in Table 7.1. The first two rows represent the total number of items recalled on the first \(T_1\) and second \(T_2\) tests, respectively, with the difference \(T_2 - T_1\) in the third row. The fact that subjects performed better on the second test than on the first test was the finding that provided so much interest to Ballard and others. After all, the second test occurs a considerable amount of time after the first test and thus one would normally expect forgetting to occur between the two tests, at least in retention of the list. However, performance on the second test was actually better than that on the first test. (No inferential statistics were applied to Brown’s data, but results like his have been frequently replicated, as the remainder of this chapter attests.)

Brown also broke down performance into two components of interest, intertest forgetting and intertest recovery. Despite the fact that more items were recalled on the second test than on the first test, some items were actually forgotten between the two tests. That is, states or words recalled on the first test were forgotten on the second test. These data are presented in the fourth row of Table 7.1, and indicate that more words than states were forgotten be-

<table>
<thead>
<tr>
<th></th>
<th>States</th>
<th>Word List</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test 1</td>
<td>36.31</td>
<td>25.48</td>
</tr>
<tr>
<td>Test 2</td>
<td>39.66</td>
<td>26.77</td>
</tr>
<tr>
<td>Difference (T_2 - T_1)</td>
<td>3.35</td>
<td>1.29</td>
</tr>
<tr>
<td>Intertest Forgetting</td>
<td>1.94</td>
<td>3.04</td>
</tr>
<tr>
<td>Intertest Recovery</td>
<td>5.29</td>
<td>4.33</td>
</tr>
</tbody>
</table>
between tests. In the fifth row is the number of items recovered between the two tests, or items not recalled on the first test that were recalled on the second. Obviously, the overall improvement between tests for states (and to a lesser extent for words) reflects the fact that intertest recovery exceeded intertest forgetting in Brown's experiments.

Ballard (1913) defined reminiscence as "the remembering again of the forgotten without re-learning" (p. v). The proper index of reminiscence then is intertest recovery—the number of items recalled on a second test that could not be recalled on a first test (with no intervening study between tests). However, Ballard (1913) occasionally examined reminiscence by reporting the overall gain between two tests, even though he clearly made the distinction between improvement between two tests and intertest recovery (pp. 17-18). Unfortunately, as Erdelyi (1984) and Payne (1987) have documented, this inconsistency in applying the term reminiscence led other researchers to the conclusion that reminiscence was not a reliable phenomenon. These investigators defined reminiscence as overall improvement between tests and noted that many experiments failed to reveal this outcome. For example, Buxton (1943) reviewed the literature and concluded that reminiscence was an unreliable phenomenon. However, the careful examination of the literature provided by Payne (1987) reveals that, when properly defined as intertest recovery, reminiscence almost always occurs in experiments involving repeated memory tests. That is, information is recalled on later tests that could not be recalled on earlier ones, even though overall improvement in the amount of information recalled between tests might not occur. Buxton's (1943) pessimistic review was probably partly responsible for the decline in theoretical and empirical interest in reminiscence until the issue was reopened by important new work by Matthew Erdelyi in the 1970s.†

Because of the historical confusion in using the term reminiscence, the current chapter will follow the relatively recent practice of defining reminiscence (as did Ballard) as intertest recovery, or recall of items on a second (or later) test that could not be recalled on a first. The overall improvement between tests will be referred to as hypermnnesia, following the reintroduction of this term by Erdelyi and Becker (1974). Hypermnnesia reflects the fact that performance increases between tests over time, in contrast to more typical forgetting or amnesia over time.

HYPERMNESIA

The resurgence of interest in hypermnnesia can be traced to important studies of Erdelyi and his colleagues (e.g., Erdelyi & Becker, 1974; Erdelyi & Klein-

†At least three other lines of research have employed repeated testing paradigms: (a) Estes' (1960) "RTT" paradigm with paired associate learning (see, for example, Izawa, 1971, 1981); (b) Tulving's (1967) repeated free recall paradigm with very limited recall time; and (c) the issue of improvements over time in motor learning (Eysenck & Frith, 1977). These literatures are only partially relevant to issues of the present chapter and will be cited when appropriate.
bard, 1978). We present the results of an experiment by Erdelyi, Finkelstein, Herrell, Miller, and Thomas (1976) because it serves to make both the relevant empirical and theoretical points most strongly. Three groups of subjects were shown different types of material with instructions to study the items carefully in preparation for a recall test. One group received 60 pictures, simple sketches of objects such as flag or trumpet. Another group received 60 words, the names of the pictures. A third group of subjects was also shown the 60 words, but in addition, were instructed to form vivid images of each named object as it was presented.

Following study, subjects were given three successive 7-minute tests at retention intervals of 1 minute, 15 minutes, and 29 minutes and told to provide as many list items as possible. (Subjects who studied pictures were told to write down names of the pictures.) A forced recall procedure was used on each test; subjects were given sheets with 40 spaces and told to write down 40 items, even if they had to guess. The forced recall procedure insured that response criteria did not change across test. (Prior work had shown that 40 responses would be well above the average number recalled.) Between tests, subjects were instructed to sit back and think about the stimulus set in preparation for the next test.

![Graph showing recall results](image)

**FIG. 7.1.** Results of Erdelyi et al. (1976). Recall of pictures and imaged words increased across tests, but recall of words studied without imagery instructions did not.
TABLE 7.2
Forgetting and Reminiscence in Erdelyi et al. (1976)

<table>
<thead>
<tr>
<th></th>
<th>Forgetting</th>
<th>Reminiscence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number¹</td>
<td>As a Proportion of Test 1 Recall</td>
</tr>
<tr>
<td>Pictures</td>
<td>5.0</td>
<td>.19</td>
</tr>
<tr>
<td>Words-Image</td>
<td>4.4</td>
<td>.17</td>
</tr>
<tr>
<td>Words</td>
<td>3.3</td>
<td>.15</td>
</tr>
</tbody>
</table>

¹Number of items recalled on Test 1 that were not recalled on either Test 2 or Test 3 (or both).
²Number of items not recalled on Test 1 but later recalled on either Test 2 or Test 3 (or both).

The results are shown in Fig. 7.1. As can be seen, recall of pictures increased substantially across the three tests. When words had been studied without imagery instructions, the total number recalled was relatively constant. However, when words had been presented with instructions to form images, recall resembled the pattern for pictures. Thus, both pictures and imaged words showed hypermnesia, whereas presentation of words with no special instructions did not. The picture/word difference replicated prior experiments by Erdelyi and Becker (1974). In response to this consistent pattern of results, Erdelyi et al. (1976) proposed that hypermnesia only occurred for information coded in memory in an imaginal format: Pictures and imaged words showed the effect, whereas recall of verbal materials did not in their experiment or in many others (but see Brown, 1923 and Table 7.1).

Further evidence for the idea that pictures and words whose referents were imagined are especially susceptible to “spontaneous recovery” after being forgotten can be found in Table 7.2, which presents intertest forgetting and reminiscence in Erdelyi et al.'s (1976) experiment. The left hand side of the table shows the number of items that were recalled on the first test, but were forgotten on Test 2, on Test 3, or on both. The number of pictures forgotten was actually somewhat greater than the number of words forgotten, with the Words-Image condition falling in between. However, because more pictures than words were recalled on Test 1, more opportunities existed for forgetting in this condition. When the forgetting figures are expressed as a proportion of the number of items forgotten after Test 1 relative to the number of items recalled during Test 1 (the second column), the differences in forgetting among types of material are less apparent. Thus, forgetting of pictures or words—given that they were recalled on Test 1—was roughly constant in the Erdelyi et al. (1976) data.²

²The forgetting data in Table 7.2 ignore one type of forgetting across three tests: items forgotten on Test 1, recalled on Test 2, but forgotten again on Test 3. The mean number of items exhibiting this pattern of forgetting in each condition was .9 for Pictures, .6 for Imagined Words, and .6 for Words studied without imagery instructions.
On the other hand, the reminiscence results at the right side of Table 7.2 exhibit a very different pattern. The third column shows the number of new items recalled after Test 1 for the three conditions. The Picture groups exhibited more spontaneous recovery than did the Words-Image group, which in turn outperformed the group that studied only Words without special instructions. These differences become even larger when taken as a proportion of the number of potential items that could be recalled after Test 1, because the pool of remaining target pictures was smaller than that for target words due to differential recall in Test 1. (That is, because 26.65 Pictures were recalled and 22.10 Words were recalled on Test 1, the pool of remaining targets was 33.35 for Pictures and 37.90 for Words. The proportions in the rightmost column reflect observed reminiscence relative to the total possible amount.) Thus, much greater reminiscence occurred for pictures and imagined words than for words, although even with words subjects recalled 9% of the targets on Tests 2 and 3 that were not produced on Test 1. In line with the findings of many others (e.g., Tulving, 1967), reminiscence occurs for words but often tends to be balanced by intertest forgetting. Pictures and imaged words show hypermnesia because the reminiscence outweighs the intertest forgetting.

The remainder of this chapter is devoted to reviewing research, largely from our laboratory, seeking to answer basic questions about reminiscence and hypermnesia. The questions include: What causes hypermnesia? Is repeated testing necessary for its occurrence? Is the forced recall procedure a necessary condition? Why is hypermnesia greater for pictures (and for imaged words) than for words? Do other variables also affect hypermnesia? What theoretical constructs are necessary for understanding it? The following sections are devoted to answering these questions, as well as to raising a few other issues.

THE ROLE OF REPEATED TESTING

One fundamental question about hypermnesia is the extent to which repeated testing is responsible for the phenomenon. This issue arose in the early work on reminiscence in which investigators sought to determine whether improvements across time could occur without repeated testing. Indeed, when repeated testing was eliminated by using between-subjects designs, the typical finding was either forgetting or no change of material over time (see McGeoch & Irion, 1952, chap. 5).

Shapiro and Erdelyi (1974) examined this issue again using Erdelyi's newly developed paradigm involving recall of pictures and words. They presented subjects with either 60 pictures or 60 words and then tested them once, either after a delay of 30 seconds or of 5 minutes. Both variables were manipulated between subjects. During the retention interval, subjects were instructed to review the material covertly. Shapiro and Erdelyi reported that people who
studied pictures performed better after 5 minutes than after 30 seconds (33.1 vs. 29.6, respectively), but there was no reliable difference in performance between the two tests for people who studied words (22.8 vs. 24.3). If taken at face value, these results would indicate that recall increased over the retention interval for pictures without practice due to testing. However, the procedure of providing subjects with a 5-minute interval to think about the material may have effectively extended the subjects’ recall time, or permitted them practice at repeatedly retrieving the items. Of course, one would then need to explain why the improvement occurred for pictures but not words, but a possible solution to this issue is described in the following section.

Roediger and Payne (1982) also addressed the issue of whether improvements in net recall across repeated tests were attributable to some change in the memory trace correlated with time (e.g., consolidation of imaginal representations) or was due to practice effects arising from repeated tests. To separate these explanations, they presented subjects with a set of 60 pictures and then gave them three successive free recall tests. Three different groups of subjects were tested, with the primary variable being the delay of the first test. The retention interval prior to the first test was filled by reading a passage from the Scientific American to prevent rehearsal or retrieval practice on the pictures. Reading was deemed unlikely to provide much retroactive interference for the pictures. Thus, for the three different groups, the first test occurred after either a Short, Intermediate, or Long delay. For subjects in the Intermediate and Long delay conditions, their first test began at the same time that the Short delay subjects began their second and third tests, respectively (see Table 7.3). The design outlined in Table 7.3 disentangles the natural confounding in repeated testing studies between retention interval and the number of prior tests.

Three facts emerge from the data presented in Table 7.3. First, by looking at the rows, note that hypermnnesia occurred in each condition, as recall improved across the three tests. (The magnitude was the same as in the Picture condition in Erdelyi et al., 1976.) Second, an examination of the downward diagonals in Table 7.3 shows that improvements in recall did not occur across

<table>
<thead>
<tr>
<th>Condition</th>
<th>Study</th>
<th>Retention Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$T_1$   $T_2$ $T_3$</td>
</tr>
<tr>
<td>Short</td>
<td>60 Pictures</td>
<td>25.6     27.9  30.1</td>
</tr>
<tr>
<td>Intermediate</td>
<td>60 Pictures</td>
<td>$T_1$ $T_2$ $T_3$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>25.1     27.5  29.8</td>
</tr>
<tr>
<td>Long</td>
<td>60 Pictures</td>
<td>$T_1$ $T_2$ $T_3$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>25.6     28.9  31.3</td>
</tr>
</tbody>
</table>
retention intervals in the absence of repeated testing. For example, subjects in the Immediate, Short, and Long delay conditions recalled 25.6, 25.1, and 25.6 items, respectively, on their initial tests. On the other hand, when the retention interval was held constant, recall increased directly with the number of prior tests. For example, in the middle column of Table 7.3, data are shown for the three groups with the retention interval held constant but the number of previous tests varied. Looking from the bottom to the top of the column (T₁, T₂, and T₃), it is obvious that recall improved directly as a function of prior tests and that the increase is the same as in the “normal” repeated testing conditions represented by data in the rows. The conclusion from Table 7.3 is that hypermnnesia for pictures depends directly on the number of prior tests and does not occur over time in absence of retrieval practice.

**CUMULATIVE RECALL AND THE RECALL LEVEL HYPOTHESIS**

At about the time Erdelyi and Becker (1974) were publishing their seminal work on hypermnnesia, the senior author was engaged in two other lines of work in which cumulative recall data were being collected (Roediger, Stellon, & Tulving, 1977; Roediger & Tulving, 1979). Cumulative recall curves portray how recall changes over time during testing and can be collected by the simple expedient of asking subjects to draw lines at regular intervals during the testing period. Because such data illuminate dynamic aspects of recall, they have been studied, sporadically, for over 40 years. For example, the data in Fig. 7.2 come from Bousfield and Sedgewick (1944), who asked subjects to write down as many items as they could from common categories during a 16-minute period. As can be seen, group cumulative recall curves are quite orderly and are well fit by an exponential equation of the following form: \( n(t) = n(\infty)(1 - e^{-\lambda t}) \), where \( n(t) \) represents the number of items that have been recalled by time \( t \), \( n(\infty) \) is the asymptote of the function (the lower bound estimate of the number of items accessible), \( e \) is the base of the natural algorithm, and \( \lambda \) is the rate of approaching the asymptote.

Surprisingly little work has been directed at understanding the dynamics of cumulative recall (although see Indow & Togano, 1970 and Vorberg & Ulrich, 1987). Our current interest lies in the use of such an analysis in understanding reminiscence and hypermnnesia. One more fact about the curves shown in Fig. 7.2 is necessary for our analysis: The parameter reflecting the rate of approaching the asymptote, \( \lambda \), is negatively correlated with the asymptotic level of recall, \( n(\infty) \). The more items one knows in a category, the slower is one’s rate of approaching the asymptotic level. This can be seen in Fig. 7.2 by comparing the cumulative recall curves from different sized categories. When recalling smaller categories, such as *unpleasant objects* or *European cities*, subjects approached the asymptote more quickly than when recalling
larger categories, such as pleasant objects, or U.S. cities. For example, if we use recall at 16 minutes as an estimate of asymptotic recall, at 10 minutes subjects recalling European cities produced 85% of the responses whereas subjects recalling fellow college students recalled 76% of their responses (and note that the latter subjects had obviously not reached asymptote). The importance of the negative correlation between $\lambda$ and $n(\infty)$ in our interpretation of hypermnnesia will become apparent.

The question that initiated our work on hypermnnesia was whether subjects given three successive tests in the standard Erdelyi and Becker (1974) paradigm would recall more information than would subjects given a single test of equivalent duration. Because cumulative recall curves often show gains even after long periods of time, as in Fig. 7.2, it seemed possible that the procedure of giving successive tests showed increases in recall simply because subjects were given additional time on later tests. Roediger and Thorpe (1978, Experiment 1) asked this question by presenting subjects either with 60 pictures or 60 words (the names of the pictures) and then testing them either with three successive 7-minute tests or one extended (21-minute) test. Free recall instructions were given in all tests and subjects were asked to draw a line after each minute of the test.

Three interesting facts emerged from this experiment. Two of these can be gleaned from results shown in Fig. 7.3, which portrays cumulative recall of the two groups of subjects given three successive tests after studying either pictures or words. First, note that hypermnnesia occurred for both pictures
and for words, although it seems somewhat greater for pictures. The usual method of portraying hypermnnesia results, such as in Fig. 7.1 of this chapter, is simply to present the number of items recalled at the end of each test—the far right hand points in Fig. 7.3. However, a second interesting fact to emerge from this analysis is that the magnitude of hypermnnesia might be greater if somewhat shorter tests were given (e.g., 3 to 4 minutes, rather than 7). Also, note from Fig. 7.3 that if only 1 minute had been given for recall, no evidence of hypermnnesia would have been obtained. Erdelyi and Becker (1974) noted that early findings of hypermnnesia might not have been well replicated because people typically employed verbal materials in their experiments. The data in

---

1This claim rests on the assumption that three 3-minute tests would reveal the same pattern of improvement as would three 7-minute tests at the end of the third minute, an untested assumption (but see Payne, 1986, Experiment 3). We would like to thank J. H. Neely for calling this to our attention.
Fig. 7.3 suggest that another consideration in obtaining the effect is the amount of recall time provided, because in many experiments this was quite brief. For example, Tulving (1967, Experiment 2) gave subjects test periods of 36 seconds to report orally 36 studied words. Although he obtained reminiscence, no hypermnnesia was observed due to great intertest forgetting. Presumably, subjects need ample time to recall prior items and new ones, as is evident in Fig. 7.3.

The third finding of Roediger and Thorpe (1978) is displayed in Fig. 7.4, where recall of pictures and words is shown for the entire 21-minute test period. For subjects given three tests, recall of an item was scored the first time it was recalled, with subsequent repeated recalls ignored. Thus, all four groups were scored by the same criterion. Although pictures were obviously recalled better than words, no difference existed in the number of unique items recalled between subjects given three successive tests and those given one long test of equivalent duration. In some sense, then, hypermnnesia occurs because subjects are simply not given "long enough" on the first test to recall all the information they know.

But why should hypermnnesia be greater for pictures than for words? Erdelyi suggested that the critical variable was whether or not information was coded in an imaginal format, but another possibility exists. Pictures are better recalled than are words studied with no special instructions, as seen in Fig.
7.1 and 7.3 and in numerous other experiments. Perhaps level of recall, rather than imaginal coding, is the relevant variable. Indeed, examination of cumulative recall curves such as those shown in Fig. 7.2 and 7.4, has suggested to us a straightforward rationale for supposing that this is so.

THE CUMULATIVE RECALL LEVEL HYPOTHESIS

The logic underlying the recall level hypothesis of hypermnesia depends on several assumptions and will be illustrated with reference to the idealized cumulative recall curves in Fig. 7.5. The following assumptions are needed to explain hypermnesia (Roediger, Payne, Gillespie, & Lean, 1982; Roediger, 1982):

1. Hypermnesia, or increased recall across repeated tests, produces equivalent total recall as compared to performance of subjects given a single test of the same duration. The results of Roediger and Thorpe (1978, Experiment 1) support this assumption and it has been repeatedly verified (see Payne, 1986, 1987).

2. Because the total number of items recalled in repeated tests is equivalent to cumulative recall, properties of cumulative recall curves may be critical for understanding the phenomenon.

3. Cumulative recall curves typically exhibit a negative correlation between the asymptotic level of recall, $n(\infty)$, and the rate of approaching that asymptote, $\lambda$ (Bousfield & Sedgewick, 1944).

![Diagram](image.png)

**Fig. 7.5.** Hypothetical cumulative recall curves to illustrate the logic of the cumulative recall level hypothesis.
4. Because the rate of approaching the asymptote is greater with lower levels of asymptotic recall, if recall is stopped prior to the asymptote, performance will be nearer the asymptotic level in cases of lower than of higher recall. This is illustrated in Fig. 7.5. Performance at the first test ($T_1$) is closer to the asymptotic level of recall in the case of the Low curve than of the High curve. To take an actual example, recall at the end of the first test in the three conditions of Erdelyi et al.'s (1976) experiment was 75% of eventual recall for subjects recalling pictures, 78% for those recalling imagined words, and 86% those recalling words without instructions (see Fig. 7.1). Thus, the higher the overall level of asymptotic recall, the greater the potential for further gains in recall after Test 1.

5. Therefore, hypermnesia (growth in recall over repeated tests) will be correlated with cumulative recall level, $n(\infty)$, or, more precisely, with the difference between cumulative recall and recall on a first test (cumulative recall $- T_1$ recall). If the logic is correct, then the reason for greater recall in one condition than another should not matter. That is, the picture/word contrast is not special in producing hypermnesia and similar results should be found by varying cumulative recall level in other ways.

6. Another assumption was noted only in passing in the original Roediger et al. (1982) formulation but it has turned out to be critical. This is the assumption that intertest forgetting does not vary with recall level. Roediger and his colleagues noted that "Differential intertest forgetting between conditions is not taken into consideration in this account" (p. 639). Of course, evidence available at the time was consistent with the assumption that intertest forgetting did not differ over conditions (Erdelyi et al., 1976, as shown in Table 7.2 here, and Roediger & Thorpe, 1978, their Table 1). However, as will be documented later, this assumption is the weakest link in the recall level hypothesis (Payne, 1986).

The cumulative recall level hypothesis of hypermnesia provided a reasonable account of most, but not all, of the extant evidence in 1982 (see Erdelyi, 1982 and Roediger, 1982 for differing opinions on this issue, however). For example, Erdelyi et al.'s (1976) finding that hypermnesia was greater for pictures than for words accords well with the formulation. In addition, when words were studied under conditions encouraging their recoding into an imaginal format, asymptotic recall increased and so did hypermnesia. Of course, the Erdelyi et al. results are similarly in accord with the hypothesis that imagery is the critical factor.

Given this situation, the obvious need was for experiments that would manipulate cumulative recall level over a wide range (but by manipulations that would not involve imagery) to see if hypermnesia still correlated with cumulative recall level. Roediger et al. (1982) reported three experiments that conformed to this general prescription and found positive results. Only Experiment 3 will be reported here, because it manipulated level of recall over the widest range. In this experiment we asked subjects to free recall natural
categories from semantic memory. We picked three categories (Presidents, Birds, and Sports) that varied considerably in size in order to induce differences in asymptotic levels of recall. Subjects were given three successive 10-minute tests, with short breaks between tests simply to collect recall sheets and pass out new ones. The instructions were to recall as many members of one category as possible each time.

The cumulative recall curves for the 30 minutes of testing are shown in Fig. 7.6 and indicate that we did succeed in manipulating asymptotic level of recall. (Indeed, recall of sports, and to a lesser extent of birds, were not at asymptote even after 30 minutes.) Also apparent from Fig. 7.6 is the fact that at the end of the first test, subjects in the various conditions had attained different proportions of their eventual level of recall. Little room for further improvement after 10 minutes existed in the condition in which subjects recalled Presidents, but subjects who recalled sports were far from asymptote after 10 minutes and therefore had considerable room for improvement. The prediction of the recall level hypothesis is simply that hypermnesia should
be a direct correlate of the difference between initial and asymptotic levels of recall, as shown in Fig. 7.5. Hypermnnesia is plotted in the customary way in Fig. 7.7, and this expectation is indeed borne out. Hypermnnesia was greater for sports than for birds, and greater for subjects recalling both these categories than for those recalling Presidents. The interaction between number of tests and the conditions was statistically significant. However, even in the case of recall of Presidents, reliable hypermnnesia was obtained, because almost all subjects showed the effect.

The experiment just reported accords well with the cumulative recall level hypothesis: Manipulation of cumulative recall level predicted hypermnnesia, and in a situation where imagery probably played little differential role in recall of the three categories (but see Erdelyi, 1982). Roediger et al. (1982) reported two other experiments in which recall was manipulated either through a levels of processing manipulation with words (Experiment 1) or by one or three repetitions of nonsense syllables (Experiment 2). The expected result was obtained in both cases, with hypermnnesia greater for those cases in which there was greater cumulative recall. Payne and Roediger (1987) obtained similar results.

**PROBLEMS FOR THE RECALL LEVEL HYPOTHESIS**

The cumulative recall level hypothesis of hypermnnesia is built on the five assumptions previously described. Erdelyi (1987) has pointed out that the recall

---

**FIG. 7.7.** Recall of Presidents, Birds, and Sports across three successive tests (Roediger et al., 1982, Experiment 3).
level hypothesis is really a statement about reminiscence, with the assumption that hypermnnesia can be directly predicted from the amount of reminiscence. He further points out that this is true only if one ignores the possibility of forgetting between tests. Indeed, the original formulation proposed by Roediger et al. (1982) did assume, almost in passing, that the amount of intertest forgetting would not differ across conditions of interest. However, this assumption turns out to be wrong in some cases, as demonstrated in experiments by Payne (1986).

Payne's (1986) experiments were designed as a test of several aspects of the cumulative recall level hypothesis. For present purposes, the first two reported experiments (Experiments 1 and 1A) are the most relevant. He presented subjects with pictures or words and manipulated the level of their recall by varying the number of presentations of the material or the amount of study time. The idea guiding Payne's work was to break the natural correlation between type of material and level of recall. When presentation conditions are held constant, pictures are better recalled than words. By manipulating the number of presentations and the amount of time, Payne sought to achieve conditions in which words would be recalled as well as or better than pictures. Assuming this condition was met, he could then determine if, under these conditions, hypermnnesia would be greater for words than for pictures.

Two examples from Payne's results illustrate his major findings. In Experiment 1A, the cumulative recall curve for words presented at a slow rate was

![Graph showing cumulative recall of pictures and words presented at different rates of presentation (Payne, 1986).](image-url)
roughly equivalent to that of pictures presented at a fast rate, as shown in
the middle curves of Fig. 7.8. Thus, by the logic of the cumulative recall level
hypothesis, one would expect equivalent hypermnesia in these conditions.
However, the results showed greater hypermnesia for pictures than for words
(Fig. 7.9, middle lines). The reason for this occurrence was that greater in-
tertest forgetting occurred for words than for pictures, thus offsetting the
equivalent amounts of reminiscence. The same pattern was obtained in another
experiment in which cumulative recall was equated by presenting words more
often than pictures. Thus, the assumption that intertest forgetting is equivalent
for different types of materials (or other conditions) was not met and conse-
quently predictions from the cumulative recall level hypothesis were not borne
out. When recall of pictures and words was equated at asymptote, pictures
were more likely to be recalled across tests than were words. Waring and Payne
(1987) provided other data that also challenge the cumulative recall level
hypothesis on similar grounds.

It is unclear why Payne (1986) found that variation in materials led to dif-
fferential intertest forgetting, but Erdelyi et al. (1976) and Roediger and Thorpe
(1978) did not. However, to the extent that intertest forgetting differs across
other independent variables that affect recall level, the cumulative recall level
hypothesis may be compromised. Payne (1986) modified the cumulative recall
level hypothesis so as to take into account intertest forgetting. Quite obviously,
the cumulative recall level hypothesis sets an upper bound on the amount
of hypermnesia that can be attained and points to factors critical in deter-
mining the phenomenon, but the problem of interest forgetting must be taken into account (Erdelyi, 1987).

Payne (1986, Table 3) reported a correlational analysis that can be interpreted as good support for the cumulative recall level hypothesis. Across subjects, he correlated (a) the amount of hypermnesia \((T_3 - T_1)\) and (b) the difference between cumulative recall and performance on the first test (Cumulative recall – \(T_1\)). This latter entity is, according to the hypothesis, responsible for hypermnesia (see Assumption 5 in the previous section). Payne (1986) found correlations of .82 and .91 for two groups of subjects recalling pictures, and correlations of .74 and .76 for two groups recalling words. All four correlation were impressive, and were higher than when hypermnesia was correlated with other values (recall on Test 1, or cumulative recall) across subjects. The higher correlations for subjects recalling pictures than those recalling words probably owes to the fact that, in Payne's (1986) experiments, less intertest forgetting occurred for pictures than for words.

Our conclusion is that the cumulative recall level hypothesis captures primary factors for predicting hypermnesia. However, it obviously founders when intertest forgetting is great, when it varies across conditions, and/or when it is negatively correlated with reminiscence. The conditions leading to such occurrences are poorly understood.

**THEORETICAL MECHANISMS EXPLAINING HYPERMNESIA**

The cumulative recall level hypothesis of hypermnesia provides a functional account of the phenomenon by identifying conditions thought to be necessary for its occurrence. However, as formulated, no molecular theoretical mechanisms are implicated. What theoretical mechanisms might cause hypermnesia? Interestingly, the two primary mechanisms usually introduced to explain hypermnesia date back to the time of Ballard (1913) and Brown (1923) and are still the most prominently mentioned today. One is the assumption that the act of retrieving and recalling information serves to make this information more likely to be accessible in the future (Erdelyi & Becker, 1974; Roediger & Thorpe, 1978). Much evidence supports this idea (e.g., Darley & Murdock, 1971), although apparently provision must be made for differential effects of prior recall on later recall for different types of material in some cases (pictures and words; Payne, 1986).

A second assumption is required to account for increased recall over time—for reminiscence or spontaneous recovery. One natural theoretical suggestion to account for reminiscence is Estes' (1955, 1959) stimulus sampling theory, which was originally applied to the similar problem of spontaneous recovery following extinction of a conditioned response. Reminiscence may be due to repeated sampling of elements from memory, with the particular elements
sampled changing within increased delays due to inherent variability in the population of elements being sampled. Stated somewhat differently, the increased recall with extended time and effort may be related to other cases where recall improves due to presentation of explicit retrieval cues (Roediger & Thorpe, 1978, p. 304). Just as provision of external retrieval cues can sometimes greatly increase recall (Tulving & Pearlstone, 1966), perhaps increased recall across time is due to changes in internal retrieval cues used by subjects (Tulving, 1974).

Direct evidence on this second assumption is less prevalent, but an experiment by Payne (1986, Experiment 4) is at least suggestive. Prior work by Tulving and Psotka (1971) and by Paris (1978) had shown that hypermnesia occurred for free recall of categorized lists. The primary question in Payne's experiment was whether or not the increased recall across repeated tests was due to (a) more different categories being added to recall, (b) more items being recalled from the same categories that had already appeared in recall, or (c) both factors. If the improved recall across repeated tests is due to the addition of novel categories during the later tests, then this would provide support for the idea that subjects are cuing themselves by thinking of category names in later recall attempts. Indeed, Payne's data showed that additional recall on repeated tests was due to increases in the number of categories recalled across tests, rather than an increase in the number of items per category recalled.

The most explicit account of hypermnesia occurs in the SAM (Search of Associative Memory) model of Raaijmakers and Shiffrin (1980, 1981). Two parameters in their model correspond to the two factors described above. "Incrementing" is the process whereby a recalled item is strengthened and made more likely to be recalled in the future; a second parameter is responsible for repeated sampling with replacement over long periods of time, at least until the same information is repeatedly retrieved and the recall process stops. Simulations have shown that both these parameters are needed within the model for hypermnesia to occur. Although the SAM model provides the best theoretical account of hypermnesia, the two critical ideas are really quite old ones and have been repeatedly mentioned as the leading candidates. Also, both ideas are embarrassingly close to redesccriptions of the relevant data, and thus we may hope for future theoretical developments that might permit clearer insight into these phenomena.

Recall Criteria and Hypermnesia

The concluding section of this chapter concerns the problem of recall criteria, which has often been neglected in the analyzing recall processes in remembering. In studies of recognition memory, researchers have worried for years over the problem of guessing and of various corrections for guessing. The application of signal detection theory to recognition memory (e.g., Bernbach, 1967)
and the use of other correction procedures have had great impact. However, the issue of differences in recall criteria across conditions has rarely been raised. Of course, the problem of recall criteria and whether or not reported memories are accurate or fabrications occurs in many situations, such as eyewitness testimony (Wells & Loftus, 1984) and hypnotic memory enhancement (Smith, 1983). Klatzky and Erdelyi (1985) provided an excellent discussion of the criterion problem in recall, considering especially recall under hypnosis.

We became interested in the problem of recall criteria through our investigations of hypermnnesia, and in particular in attempting to determine why some researchers find little evidence for hypermnnesia for words, whereas our studies have routinely shown this effect (cf. Erdelyi et al., 1976; Roediger & Thorpe, 1978). One possible reason for the difference is that we have typically employed free recall with instructions warning subjects against guessing, whereas Erdelyi and others have employed a forced recall procedure in which subjects are instructed to produce a fixed number of responses, guessing when necessary. Presumably subjects are much more conservative in responding when engaged in free recall than in forced recall, as the low intrusion rates in free recall attest. The forced recall procedure insures that response criteria cannot change across successive tests, but this possibility remains open when free recall is employed. Perhaps hypermnnesia occurs for words in free recall, but not forced recall, because subjects loosen their recall criteria across the tests in the former case and are more willing to guess on later tests.

In order to test this possibility, Roediger and Payne (1985) had subjects study 70 words and then take three successive tests. However, different groups of subjects received tests under one of three instructional conditions. The free recall group received typical free recall instructions with a warning against guessing; the forced recall group was told to write down at least 40 responses on each test; a third group, referred to as the uninhibited free recall condition (after Bousfield & Rosner, 1970) was told to try to remember the words as well as possible, but to report any items that came to mind during the test phase, whether or not they deemed the retrieved responses correct.

This instructional manipulation had a huge effect on the intrusion rate across the 24 minutes of testing (three 8-minute tests), as can be seen in Table 7.4. Free recall subjects intruded 7.5 items, uninhibited recall subjects intruded 28.8, and forced recall subjects intruded 71.4 across 24 minutes. (The rates in Table 7.4 are means of the three tests.) However, despite this huge difference in recall criteria as reflected by the intrusion rates, hypermnnesia was the same in all three conditions, as can be seen in Table 7.4. Thus, the differences in testing procedure do not explain the variable outcomes in obtaining hypermnnesia for words. However, of more interest in the present context is the fact that overall recall levels did not differ despite wide variations in recall criteria. That is, even though subjects in the forced recall condition produced many more intrusions than did free recall subjects, the total number of words correctly recalled did not differ. Across the 24-minute testing period, free recall subjects produced 31.2 target
TABLE 7.4
Results From Roediger and Payne (1985)

<table>
<thead>
<tr>
<th>Instructions</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>Mean</th>
<th>Intrusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Free Recall</td>
<td>25.1</td>
<td>26.6</td>
<td>28.2</td>
<td>26.6</td>
<td>2.5</td>
</tr>
<tr>
<td>Uninhibited Recall</td>
<td>24.3</td>
<td>26.4</td>
<td>28.4</td>
<td>26.4</td>
<td>9.6</td>
</tr>
<tr>
<td>Forced Recall</td>
<td>25.1</td>
<td>26.0</td>
<td>27.4</td>
<td>26.2</td>
<td>23.8</td>
</tr>
<tr>
<td>Mean</td>
<td>24.8</td>
<td>26.3</td>
<td>28.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

items whereas the figures for the uninhibited recall and forced recall conditions were 31.0 and 30.9, respectively. Surprisingly, varying recall criteria had no effect on the total number of words correctly recalled.

Roediger and Challis (in preparation) performed several other experiments exploring this unexpected effect. These experiments have led to a further puzzle, which we do not yet understand. However, we will present selected data from one experiment to illustrate the problem. Four groups of subjects in this experiment studied 60 pictures and then received either free recall or forced recall tests. Two groups of subjects received two tests, one relatively soon after presentation of the material (Immediate) and then the same type of test again after a 1 week delay (Delayed). Subjects were given 10 minutes to recall the material, with forced recall subjects being required to write down 60 items (and encouraged to write more if they so desired). Free recall subjects were cautioned against guessing, but asked to continue trying during the entire testing period. Two other groups of subjects received the same set of pictures, but were only tested after a one week delay. Following each test, subjects were asked to give confidence ratings as to how likely each recalled item was to have actually occurred in the study list. Subjects rated each item on a 6-point scale, where 1 indicated that subjects were highly confident that the items had not appeared on the study list and 6 indicated high confidence that the item had appeared on the study list.

Results are shown in Table 7.5. The top two rows contain performance of the two groups tested twice. Notice that on the immediate test, the forced recall subjects produced more correct responses than did the free recall subjects (a significant difference). However, when performance was conditionalized only on those items subjects believed to have actually appeared on the study list (4, 5, and 6 responses on the confidence scale), the number of items recalled did not differ between free and forced recall conditions. This occurred despite the fact that the intrusion rates differed substantially, as shown in the third column. This pattern of results was replicated when these subjects were tested a week later, although overall performance was of course lower; again forced recall subjects produced significantly more correct responses during the test, but this difference disappeared when guesses were eliminated.

A different pattern of results was obtained in the two groups that were
TABLE 7.5
Results From Roediger and Challis (in preparation)

<table>
<thead>
<tr>
<th></th>
<th>Test 1 (Immediate)</th>
<th>Test 2 (One Week Delay)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Correct</td>
<td>High Conf.*</td>
</tr>
<tr>
<td>Free</td>
<td>29.2</td>
<td>28.2</td>
</tr>
<tr>
<td>Forced</td>
<td>32.8</td>
<td>28.8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Delayed Test Only</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Correct</td>
</tr>
<tr>
<td>Free</td>
<td>16.7</td>
</tr>
<tr>
<td>Forced</td>
<td>25.9</td>
</tr>
</tbody>
</table>

*The number of correctly recalled items that subjects later judged to have been from the study list.

**The number of intrusions.

only given a delayed test (see the bottom of Table 7.5). Forced recall subjects again produced more correct responses than did free recall subjects, but this difference survived (in attenuated form) even when guesses were eliminated and credit was given only for responses that subjects judged to be correct. We generally replicated this pattern under other conditions of presentation: Variation in recall criteria appears to have an effect on recall on delayed, but not on immediate, tests.

The theoretical interpretation of these results is unclear, to say the least. However, two points can be made. First, unlike the Roediger and Payne (1985) experiments, we did find an effect of recall criteria in recall of pictures (see also Erdelyi, Finks, & Feigen-Pfau, in press). Forced recall subjects produced more correct responses than did free recall subjects, although the differences in “hit rates” seems relatively slight when compared to the huge difference in intrusions (or the “false alarm” rates). Perhaps the concrete and higher frequency nature of the materials used in this experiment—because we had to use items that could be converted into recognizable sketches—permitted better guessing. The second point is that, when subjects provided confidence ratings, the effect of varying recall criteria either disappeared (top of Table 7.5) or shrivered (bottom of Table 7.5). Thus, even when subjects produced more correct items under forced recall conditions, they were largely unaware of this benefit. This may represent a type of free association priming as studied in implicit memory tests (see Schacter, 1987). The unravelling of these puzzles must await further research.

CONCLUDING COMMENTS

This chapter has delineated one approach to the problems of reminiscence and hypermnnesia. The cumulative recall level hypothesis has the advantage
of (a) providing an explicit (if functional) formulation of these phenomena, (b) stating and bolstering its six primary assumptions, (c) providing a testable account, and (d) drawing attention to the dynamics of recall over time. Although the cumulative recall hypothesis can account for a number of findings in the hypermnesia literature, it founders on the assumption of equivalent intertest forgetting across conditions. Whenever conditions differ in intertest forgetting, the recall level hypothesis may not accurately predict hypermnesia (Payne, 1986). However, if differences in intertest forgetting are slight, and if levels of recall are varied over a large range (as in Roediger et al. (1982), Experiment 3; see Fig. 7.5 and 7.6), then the recall level hypothesis may survive despite the invalidity of the intertest forgetting assumption. Future work will likely center on development of theoretical mechanisms to explain hypermnesia, although the SAM model (Raaijmakers & Shiffrin, 1980) provides a reasonable account. In addition, another profitable direction for future research will be the exploration of how differences in recall criteria affect remembering.

ACKNOWLEDGMENT

Research reported in this chapter was supported by Grant R01 HD15054 from the National Institute of Child Health and Human Development.

REFERENCES


