THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY, 2003, *56A* (?), 000–000

# Remembering the forgotten? Reminiscence, hypermnesia and memory for order

Matthew R. Kelley
*Lake Forest College, Illinois, USA*
James S. Nairne
*Purdue University, Lafayette, Indiana, USA*

Three experiments established that repeated testing affects item and order retention differently: Hypermnesia was found with repeated free recall tests, whereas net performance declined significantly across successive free reconstruction of order tests. Overall order performance declined over tests under a variety of encoding conditions (pictures, words, and relational and item-specific processing) and retrieval conditions (intentional and incidental learning). Although net performance dropped across tests, participants did show reliable order recovery (reminiscence) between tests. The implications of these data for general theories of hypermnesia and order are discussed.

Often in our lives, we are asked a question that we are unable to answer correctly at that exact moment (e.g., who was your third grade teacher?) but, when queried a short time later, we are able to successfully answer the question (Ms. Calkins). This everyday phenomenon is termed *reminiscence* and can be defined as "the remembering again of the forgotten without relearning" (Ballard, 1913, p. v) or as remembering information on a later test that could not be remembered on an earlier test. *Hypermnesia* is related to reminiscence and is defined as an overall improvement in performance across tests. Hypermnesia occurs when more information is newly recovered on a later test than is forgotten between those tests—that is, when intertest recovery (reminiscence) exceeds intertest forgetting. It is important to note that reminiscence (recovery) is a prerequisite for hypermnesia (overall improvement in performance) but hypermnesia is not necessary for reminiscence.

Research on reminiscence and hypermnesia has employed free recall, cued recall, and recognition to measure retention (e.g., Payne, 1987; Roediger & Challis, 1989). Generally, one finds robust reminiscence and hypermnesia in free recall (e.g., Erdelyi & Becker, 1974;

Requests for reprints should be sent to Dr Matthew R. Kelley, Department of Psychology, Lake Forest College, Lake Forest, IL 60045, USA. Email: kelley@lfc.edu

The authors would also like to thank Ian Neath, Robert Proctor, E. John Capaldi, Aimee Surprenant, Mike Page, Mark McDaniel, and Janet Vousden for their helpful comments and suggestions on previous drafts of this article. We would also like to thank Michael Canavan and Ashley Drew for their hard work on this project.


http://www.tandf.co.uk/journals/pp/02724987.html DOI:10.1080/02724980244000530

McDaniel, Moore, & Whiteman, 1998; Roediger & Payne, 1982) and cued recall (e.g., Otani & Hodge, 1991; Otani & Whiteman, 1994; Payne, Hembrooke, & Anastasi, 1993), whereas repeated recognition tests typically yield reminiscence with either no change or a decline in overall performance across tests (e.g., Otani & Hodge, 1991; Otani & Stimson, 1994; Payne & Roediger, 1987). These studies have focused primarily on the patterns of item loss (e.g., how many items were forgotten between the first and the second tests?) and item recovery across tests (e.g., how many previously unrecalled items were recovered on the second test?).

Memory researchers commonly distinguish between the types of information required by mnemonic tests. One classic distinction is made between item information (what items occurred on a list) and order information (the ordinal location or serial position of an item in a list; Bjork & Healy, 1974; Healy, 1974; Murdock, 1976; Murdock & vom Saal, 1967).[1] The memory tests that investigators use to assess retention often differ in the types of information that they require. For example, item but not order information is usually important in a recognition test, whereas both item and order are potentially important in free and cued recall. To date, studies of reminiscence and hypermnesia have focused exclusively on the recovery and loss of item information and have ignored the effect of repeated testing on memory for order information (e.g., order loss and order recovery).

The experiments reported in this article were designed to establish how repeated testing affects memory for order. Free reconstruction of order was used to assess retention in these experiments. In a free reconstruction task, the items that occurred on the just-presented list are provided at the point of test, and the participant's task is simply to place the items back into their original order of presentation. Although not a pure test of order (Nairne & Kelley, 1999; Neath, 1997), reconstruction is generally regarded as a better measure of order retention than are other tasks, such as serial recall, because item information is provided at the time of testing. Thus, reconstruction performance should be driven primarily by memory for order information.

Understanding how repeated testing affects memory for order is important for a number of reasons. First, it is important to determine whether or not repeated testing has similar effects on item and order retention because dissociative effects of item and order are common in the memory literature (e.g., Bjork & Healy, 1974; Healy, 1974; Murdock, 1976). A dissociation between item and order in this situation might have significant implications for current theories of hypermnesia, which were developed with only item information in mind. Moreover, the order data provide a critical test for general theories of order memory. A hypermnesic effect in order might be difficult for current theories to handle because most theories of order do not have a mechanism that allows performance to improve across tests.

The repeated reconstruction test data are also potentially important because they allow one to consider whether the processes involved in reconstruction tap processes similar to those employed in recognition and/or recall (e.g., see Whiteman, Nairne, & Serra, 1994). For example, one could argue that a recall-like process guides performance in a reconstruction task. When participants are asked to reconstruct the order of a list of items, they might simply recall the list items in order and record them in their appropriate positions. In this case, one might expect repeated reconstruction tests to produce both reminiscence and hypermnesia, just as

[1]For the purposes of this article, order information is treated as synonymous with memory for position.

recall does. On the other hand, one could consider reconstruction to be nothing more than a fine-grained recognition task (Crowder, 1976). In recognition, participants are given an item and are asked to make a coarse-grained judgement about the item (did the item appear on the list?), whereas in a reconstruction test, participants are given an item and are asked to make a more specific judgement (where did the item appear in the list?). In this case, one might expect repeated reconstruction tests to produce reminiscence without hypermnesia, just as recognition does (e.g., Otani & Hodge, 1991).

## Repeated testing and memory for order

In a typical repeated testing experiment, participants are instructed to remember a list of items (e.g., pictures, words) for a subsequent recall test. Researchers often use a large pool of items (e.g., 40–60 items) to avoid ceiling effects, thus allowing room for performance to improve. Following the initial test, participants are surprised with a series of two or more recall tests. The tests are successive, and each lasts for about 7 min. This method has produced robust hypermnesic effects across a wide variety of designs and materials in recall (e.g., see Payne, 1987; Roediger & Challis, 1989) and has formed the backbone of the current set of experiments. Of course, a few modifications to this method were required before we were able to address the question of central interest—what effect does repeated testing have on memory for order?

First, as indicated earlier, free reconstruction of order was used to assess retention. Next, the current experiments used a variation of a procedure employed by Nairne (1990, 1991) in his work on long-term memory for order. Participants viewed a set of 25 items that were arranged into five lists of 5 items. Following presentation of the 25 items, three successive reconstruction tests were given. The tests contained five rows of five blanks corresponding to each possible list and within-list position, respectively. At the bottom of the page, the 25 items were presented in a new random order. The participants' task was to place each item back into its correct absolute position (original position within original list). Nairne (1990, 1991) already has shown that participants can complete this task with a fair degree of success; performance is commonly well above chance and well below ceiling. Also, this method provides order data along multiple dimensions. For example, one can examine how well participants are able to reconstruct the absolute order of items (e.g., correct list and within-list position) as well as how well participants fare along each dimension separately (list and within list).

This basic procedure was employed in each of the following three experiments.[2] The overall intent of these experiments was to establish how repeated testing affects memory for order under a range of conditions commonly employed in the hypermnesia literature (e.g., manipulate pictures vs. words; retention interval; relational/item-specific processing). In Experiment 1, we examined participants' performance following three successive recall or reconstruction tests. The second experiment extended the analysis by manipulating the duration of the retention interval that separated stimulus presentation and the initial test. Finally, in the third experiment, we further explored the role of encoding processes (i.e., relational and item-specific processing) in repeated-order tests.

[2]With the exception of Experiment 3, the current experiments differed from those reported by Nairne (1990, 1991) in that Nairne used incidental learning in his experiments.

# EXPERIMENT 1

In the first experiment, participants viewed a set of either 25 words or 25 pictures. Following the final item in the set, participants received three consecutive recall or reconstruction tests, depending on their respective condition. Previous research suggests that both the picture and the word recall conditions should produce hypermnesia in this task. Historically, hypermnesia is more readily obtained with pictorial stimuli than with verbal stimuli (e.g., see Payne, 1987), however, researchers have provided numerous demonstrations of hypermnesia with verbal materials (e.g., McDaniel et al., 1998; Payne, 1986, 1987). Thus, hypermnesia is expected in both picture and word recall. With regard to the reconstruction condition, it is not clear how the participants will perform. If recall-like processes are at work in reconstruction, then one might expect reminiscence and hypermnesia following repeated reconstruction tests. However, if recognition-like processes guide reconstruction, then one might expect participants to produce reminiscence without hypermnesia in this task.

## Method

### *Participants and apparatus*

Participants were 180 Purdue University undergraduates who participated for course credit in an introductory psychology course. Groups of 10 or fewer participants were tested in sessions lasting approximately 40 minutes. Stimuli were presented and controlled with an IBM-compatible computer.

### *Materials and design*

Participants were assigned to one of four experimental conditions (picture recall, word recall, picture reconstruction, or word reconstruction) based on their order of arrival to the experiment. In all, there were 45 participants assigned to each of the four conditions. In each condition, participants viewed one set of 25 items, which was arranged into five lists of 5 items. To ensure generality, three different sets of 25 items were used in each condition. A total of 75 names of common objects were taken from the Snodgrass and Vanderwart (1980) norms, and pictures of these objects were selected from Microsoft ClipArt and Blast Software's 10,000 Graphics Pack, Volume 1. These items were randomly assigned to three different sets of 25 items; the items in each set were presented in exactly the same order to all participants.

### *Procedure*

Participants were asked to remember a set of 25 items (pictures or words) for a subsequent memory test (free recall or free reconstruction). They were instructed about the exact nature of the memory test prior to list presentation, but they only expected one test. The stimuli were projected onto a screen at a rate of one item every 2 s, with a 5-s blank interval separating each of the five lists. Participants were instructed to read silently (or name) each item as it was presented. Following the final item, participants were read a brief set of instructions (for 2 min) and then received the first test.

Participants in the reconstruction condition were given a test sheet containing five rows of five blanks listed next to the numbers 1–5 (see Appendix for a sample reconstruction test sheet with stimuli employed in the experiment). They were told that the numbers corresponded to each list (e.g., 1 = list one, etc.) and the blanks corresponded to the positions within the list (e.g., first slot = first position, etc.). At the bottom of the page, the 25 words from the just-presented list were listed in three rows of 9, 8, and 8 items. Participants were asked to place each of the words listed at the bottom of the page into both its proper list and its proper position within the list, being careful not to use the same word twice. Participants in the recall condition were given a similar test sheet with two exceptions: A number between 1 and 25 appeared next to each blank, and the words from the just-presented list were omitted from the bottom of the sheet. The participants were asked to write down the names of as many pictures or words as they could remember, in any order they wished. Everyone was given 7 min to complete their test and was encouraged to use the entire time allowed to recall and reconstruct the items as accurately as possible.

Following the first test, the participants' test sheets were collected. Participants were read a brief set of instructions (2 min) in which they were asked to complete a second test, identical to the first. They were encouraged to try to improve their performance on the second test relative to the first test. The second test began immediately following the instructions and lasted for 7 min. Upon completion of Test 2, participants were given more instructions (2 min) and the third (final) test; the general procedure of Test 3 was similar to that of Test 2. Again, participants were encouraged to improve their performance on this test relative to the first two.

## Results and discussion

Table 1 displays the net recall and reconstruction levels for each test (top panel) and the rates of intertest recovery and forgetting between early and late tests (bottom panel). In the recall condition, an item was scored as correct if it appeared in the original set of items. A lenient scoring criterion was used in this condition because participants did not always use the same label to describe a picture (or word) as that intended by the experimenter (e.g., say "clock" for a picture of a "watch"); these responses were marked as correct. Such instances were rare, having occurred for only 5 of the 90 participants in the recall conditions. In the reconstruction condition, an item was scored as correct if it was placed into its original absolute position within the set of 25 items (i.e., proper position within proper list).

Two separate 2 (stimuli: pictures vs. words) × 3 (test number: 1, 2, 3) mixed-factor analysis of variance (ANOVA) tests were performed on the mean proportion of correctly recalled and reconstructed items, respectively. As expected, the recall ANOVA revealed main effects of stimuli, $F(1, 88) = 13.37$, $MSE = 0.05$, $p < .001$, and test number, $F(2, 176) = 15.66$, $MSE =$

TABLE 1
Net recall and reconstruction levels and rates of intertest recovery and forgetting in Experiment 1

| *Dependent measure* | *Test* | *Recall* | | *Reconstruction* | |
|---|---|---|---|---|---|
| | | *Pictures* | *Words* | *Pictures* | *Words* |
| Net performance | 1 | .61 | .51 | .39 | .32 |
| | 2 | .62 | .52 | .37 | .30 |
| | 3 | .65 | .55 | .36 | .28 |
| | Change (T3 – T1) | +.04 | +.04 | –.03 | –.04 |
| Recovery | Early (T1:T2) | .05 | .06 | .04 | .04 |
| | Late (T2:T3) | .05 | .06 | .05 | .04 |
| Forgetting | Early (T1:T2) | .04 | .05 | .06 | .07 |
| | Late (T2:T3) | .02 | .03 | .05 | .06 |

0.003, $p < .001$, and no interaction between these variables, $F < 1$. The recall data showed a classic picture superiority effect: Pictures were remembered better than words (e.g., see Paivio, 1974; Paivio & Csapo, 1974). In addition, importantly, the data revealed significant hypermnesia across tests for both pictures and words. Participants improved their performance by an average of 4% between the first and last tests. Although numerically small, the magnitude of this increase is consistent with previous hypermnesia research (e.g., see Payne, 1986; Payne & Roediger, 1987), and the individual participant data show this effect to be robust. Of the 90 participants in these two conditions, 60 showed hypermnesia across tests, 13 showed the opposite pattern, and there were 17 ties. These recall results are important because they confirm that the materials and design of the experiment were sensitive enough to produce hypermnesia when it should be produced—with recall. The next question is whether these materials and design were sufficient to produce hypermnesia in a reconstruction of order task.

The reconstruction ANOVA revealed main effects of stimuli, $F(1, 88) = 4.47$, $MSE = 0.08$, $p < .05$, and test number, $F(2, 176) = 7.05$, $MSE = 0.005$, $p < .001$, and no interaction between these variables, $F < 1$. Interestingly, as in the recall condition, there was a reliable picture superiority effect: Pictures were reordered better than words. To our knowledge, this is the first demonstration of picture superiority effect in a task that purportedly measures memory for order. Unlike recall, however, net reconstruction performance declined significantly across tests, by 3% and 4% for pictures and words, respectively. In other words, participants did not show hypermnesia in reconstruction; they tended to forget more than they remembered between tests. Overall performance dropped for 45 of the 90 participants in these conditions, 26 showed the opposite pattern, and there were 19 ties. Clearly, repeated testing affected item and order retention differently: It hurt the participants' ability to reconstruct the order of the items, but helped the participants remember more items in the recall condition. A 2 (test type: recall vs. reconstruction) × 2 (stimuli: pictures vs. words) × 3 (test number: 1, 2, 3) mixed-factor ANOVA confirmed this Test Type × Test Number interaction, $F(2, 352) = 19.76$, $MSE = 0.004$, $p < .001$.

Turning to the forgetting and recovery data (bottom panel of Table 1), it is important to note the presence of reminiscence (recovery) in the reconstruction of order task—participants were able to remember previously forgotten order information. The value for intertest recovery was calculated for each participant by dividing the number of newly recovered or reconstructed items on a given test by the total number of items in the list (25); intertest forgetting was determined in a similar fashion—the number of items that were not recovered or reconstructed correctly on a given test, despite having been correct on the prior test, were divided by the total number of items in the list (25). Intertest recovery and forgetting remained relatively constant across early and late tests (4% and 6% for recovery and forgetting, respectively), and the forgetting always exceeded the recovery, which led to the overall drop in performance across tests. Typically, one might have expected forgetting to decline on the late tests because of retrieval practice or because of the negatively accelerated nature of the forgetting function. Although this was not the case here, it was the case in nearly all conditions of the remaining experiments when testing occurred immediately after list presentation.

*Multidimensional order analyses.* As mentioned previously, one advantage of the design of this experiment is that order data can be gathered along multiple dimensions. In addition to examining how well participants could reconstruct the absolute order of items, one can also explore how well participants fared along list and within-list dimensions. To accomplish this, the reconstruction data were scored with respect to two separate criteria: correct list and correct within-list position. With the correct list criterion, an item was scored as correct if it was placed into its appropriate list (1–5), regardless of whether it was placed into its proper position within that list. The opposite was true for the correct within-list position criterion; an item was scored as correct if it was placed in its proper within-list position (1–5), regardless of whether it appeared in its appropriate list.

The data from the multidimensional analysis are displayed in Table 2 as a function of dimension (list vs. within list), serial position (1–5), and test number (1–3).[3] Two separate 3 (test number) × 5 (serial position) ANOVAs were performed on the mean proportion of correctly reconstructed items for the list and the within-list position dimensions, respectively. The list dimension ANOVA revealed a reliable main effect of serial position, $F(4, 356) = 56.33$, $MSE = 0.11$, $p < .001$, a marginally significant main effect of test number, $F(2, 178) = 3.02$, $MSE = 0.03$, $p < .052$, and no interaction between these variables, $F < 1$. The list serial position functions showed clear primacy and recency effects, which reflect the fact that participants were more likely to place an item in its proper list if the item originally occurred in either the first or the last list. Overall performance declined across tests by 3% and this decline appeared to be distributed uniformly across the lists. The amounts of forgetting and recovery between tests (.05 and .03, respectively) were comparable to those reported in Table 1.

The within-list position ANOVA also revealed a reliable main effect of serial position, $F(4, 356) = 41.24$, $MSE = 0.06$, $p < .001$, a marginally significant main effect of test number, $F(2, 178) = 3.01$, $MSE = 0.02$, $p < .052$, and no interaction between these variables, $F < 1$. The serial position functions were bow-shaped and showed marked primacy and recency effects. There was an overall drop in performance across tests of 3% and, as with the list dimension,

TABLE 2
Serial position functions[a] in Experiment 1

| | | *Serial position* | | | | | |
|---|---|---|---|---|---|---|---|
| *Dimension* | *Test* | *1* | *2* | *3* | *4* | *5* | *Mean* |
| List | 1 | .75 | .46 | .39 | .40 | .54 | .51 |
| | 2 | .75 | .45 | .36 | .40 | .52 | .49 |
| | 3 | .73 | .45 | .36 | .38 | .49 | .48 |
| | Change (T3 – T1) | –.02 | –.01 | –.03 | –.02 | –.05 | –.03 |
| Within-list | 1 | .68 | .53 | .44 | .44 | .55 | .53 |
| | 2 | .66 | .51 | .43 | .43 | .50 | .50 |
| | 3 | .66 | .49 | .41 | .44 | .53 | .50 |
| | Change (T3 – T1) | –.02 | –.04 | –.03 | .00 | –.02 | –.03 |

[a]Collapsed across stimulus condition.

[3]Aside from an overall main effect of stimulus condition (pictures vs. words), there were no detectable differences in the list and within-list position data patterns for the pictures and words. For this reason, and for the sake of simplicity, the data in Table 2 are collapsed across stimulus condition.

this decline was distributed evenly across the serial positions. As in the list condition, the amounts of intertest forgetting and recovery (.06 and .04, respectively) mimicked those reported in Table 1.

The goals of this analysis were to (1) determine how order performance along the list and within-list position dimensions changed across tests and (2) establish where these changes occurred (e.g., did performance decline for middle list positions only? middle lists only?) The analyses showed that reconstruction performance declined across tests (intertest forgetting exceeded intertest recovery) for both the list and the within-list position dimensions. Moreover, the drop in performance appeared to be spread evenly across lists and within-list positions.

*Summary.* Experiment 1 revealed a number of important findings. First, it established that repeated testing affects item and order retention differently: Participants in the recall condition produced hypermnesia across tests, whereas performance declined over tests for participants in the reconstruction condition. Second, participants showed reminiscence in an order retention task, although the amount of recovery tended to be less than the amount of forgetting between tests; intertest recovery and forgetting remained relatively constant over the early and late tests. Finally, the multidimensional order analyses showed that reconstruction performance dropped across tests for both the list and the within-list position dimensions, and this decline was distributed evenly across list and within-list positions.

## EXPERIMENT 2

Experiment 2 was designed to determine whether the decline in order performance across tests in Experiment 1 was the product of the increased retention interval associated with repeated testing or whether it was a function of the tests themselves. Roediger and Payne (1982) asked a similar question with regard to hypermnesia in recall. They noted that retention interval and number of tests are naturally confounded in most hypermnesia studies. Consequently, they were unable to determine whether hypermnesia was due to retrieval practice from prior tests or to changes in the strength of memory traces during the retention interval. To remedy this, Roediger and Payne (1982) developed a procedure in which one could examine the effects of retention interval and repeated testing simultaneously.

In their procedure, everyone received three successive tests, but the onset of the first test differed across participants; some people received their first test immediately, whereas others received the test after delays of 9 or 18 min. With this method, they were able to examine the effect of the number of prior tests after a constant delay, as well as the effect of retention interval after a constant amount of tests. Roediger and Payne (1982) revealed that recall performance varied with the number of prior tests, but not with retention interval. They concluded that hypermnesia in recall depends directly on repeated testing and does not occur over time without retrieval practice (see also, Roediger & Challis, 1989).

At this point, it is not clear whether the decline in order performance across tests in Experiment 1 was due to an increased retention interval or to repeated testing because these factors were confounded in the experiment. In the second experiment, retention interval and number of tests were varied simultaneously, using a variation of the Roediger and Payne (1982) method. In the experiment, participants were asked to remember a set of 25 items for a subsequent reconstruction test. Some people received the test immediately following list presentation, while others received either 9 or 18 min of distracting activity prior to the test. Upon completion of the first test, everyone was given two additional reconstruction tests, successively.

## Method

### Participants and apparatus

A total of 270 Purdue University undergraduates participated for course credit in an introductory psychology course. Groups of 10 or fewer participants were tested in sessions lasting approximately 60 min. Stimuli were presented and controlled with an IBM-compatible computer.

### Materials and design

The materials from Experiment 1 were used again in Experiment 2. The design was the same as that of the first experiment, with two exceptions. Participants were assigned to one of three delay conditions (immediate, short delay, and long delay) and one of two stimulus conditions (pictures vs. words) based on their order of arrival to the experiment; in all, there were 45 participants assigned to each of the six conditions. Also, all participants received the reconstruction of order test.

### Procedure

The procedure of Experiment 2 matched that of Experiment 1, with the following exceptions. After presentation of the final item, participants either received a brief set of instructions (for 2 min) and their first reconstruction test (immediate condition) or were given 9 or 18 min of distracting activity (short and long delay conditions, respectively). During the distraction period, participants were asked to solve mathematical problems. Following 9 or 18 min of distraction, participants in the short and long delay conditions received instructions (2 min) and their first test. All participants were given 7 min to complete the first test.

After the first test, the procedure was the same for all participants and mimicked that of Experiment 1. Participants were read a brief set of instructions (2 min) in which they were asked to complete a second test, identical to the first. They were encouraged to try to improve their performance on the second test relative to the first test. The second test began immediately following the instructions and lasted for 7 min. Upon completion of Test 2, participants were given more instructions (2 min) and a third (final) test.

## Results and discussion

Table 3 displays the net reconstruction levels on the three tests following each delay condition. As in Experiment 1, an item was scored as correct if it was placed into its original absolute position within the set of 25 items (i.e., proper position within proper list). The question of central interest in this experiment was whether the decline in order performance across tests was the product of the increased retention interval associated with repeated testing, the repeated tests themselves, or some combination of the two. To answer this question, two separate 3 (delay: immediate, short, or long) × 3 (test number: 1, 2, 3) mixed-factor analyses of variance (ANOVA) tests were performed on the mean number of correctly reconstructed items for the picture and word conditions, respectively. In addition, Newman–Keuls tests were employed so that performance could be compared for each test following each delay.

TABLE 3
Proportion correct on the tests in Experiment 2

| | *Proportion correct* | | | | |
|---|---|---|---|---|---|
| *Condition* | *A* | *B* | *C* | *D* | *E* |
| | | | Pictures | | |
| Immediate | Test 1 | Test 2 | Test 3 | | |
| | .39 | .33 | .34 | | |
| Short | | Test 1 | Test 2 | Test 3 | |
| | | .30 | .29 | .29 | |
| Long | | | Test 1 | Test 2 | Test 3 |
| | | | .31 | .29 | .30 |
| | | | Words | | |
| Immediate | Test 1 | Test 2 | Test 3 | | |
| | .34 | .31 | .31 | | |
| Short | | Test 1 | Test 2 | Test 3 | |
| | | .27 | .25 | .25 | |
| Long | | | Test 1 | Test 2 | Test 3 |
| | | | .24 | .24 | .24 |

The picture ANOVA revealed a main effect of test number, $F(2, 264) = 7.33$, $MSE = 0.004$, $p < .001$, and a Delay × Test Number interaction, $F(4, 264) = 2.42$, $MSE = 0.004$, $p < .05$, the main effect of delay was not reliable, $F(2, 132) = 2.05$, $MSE = 0.07$, $p > .13$. The Newman–Keuls test uncovered three important findings. First, the immediate, short delay, and long delay conditions produced different patterns of reconstruction performance across tests: Net performance declined significantly over tests in the immediate condition (T1 > T2 = T3), but did not change over tests in the short and long delay conditions (T1 = T2 = T3, following both delays). Of the 45 participants in the immediate condition, 26 showed an overall drop in performance, 12 showed the opposite pattern, and there were 7 ties; these results replicate the finding reported in Experiment 1. In contrast, performance in the short and long delay conditions seemed to reach asymptote after about 9 min; performance dropped for 37 of the 90 participants, while 35 showed the opposite effect, and there were 18 ties.

Second, the Newman–Keuls test revealed a reliable effect of retention interval. This effect can be seen by holding test number constant and varying the delay condition (the upper left to lower right diagonals of Table 3). Net reconstruction performance in the immediate condition was significantly higher than that in the short delay and long delay conditions following each test (Test 1: .39 > .30 = .31; Test 2: .33 > .29 = .29; Test 3: .34 > .29 = .30). This evidence suggests that participants' ability to reconstruct the order of events declines, up to a point, as the length of the retention interval increases.

Finally, the analysis showed an effect of repeated testing. This can be seen when retention interval is held constant, and the number of prior tests is varied (the columns of Table 3: A, B, C, D, and E). Column C of Table 3 shows how well participants performed on their first test (.31), second test (.29), and third test (.34) following a constant 20-min retention interval; performance was reliably better following two prior tests than it was following either one or no prior tests (.34 > .29 = .31, respectively). In other words, repeated testing had a positive effect

TABLE 4
Rates of intertest recovery and forgetting in Experiment 2

| Dependent measure | | *Pictures* | | | *Words* | | |
|---|---|---|---|---|---|---|---|
| | | *Imm.* | *Short* | *Long* | *Imm.* | *Short* | *Long* |
| Recovery | Early (T1:T2) | .04 | .05 | .05 | .05 | .04 | .04 |
| | Late (T2:T3) | .05 | .04 | .05 | .05 | .04 | .04 |
| Forgetting | Early (T1:T2) | .10 | .06 | .07 | .08 | .06 | .04 |
| | Late (T2:T3) | .04 | .04 | .03 | .04 | .03 | .04 |

on order retention by attenuating the net amount of forgetting across tests.[4] This held true for participants in the immediate condition (see also, column B: .33 > .30), but not for those in the short and long delay conditions—there was no benefit of number of prior tests in these settings (see also, column D: .29 = .29).

The data in the word condition were similar to those reported in the picture condition. The word ANOVA revealed a reliable main effect of test number, $F(2, 264) = 4.64$, $MSE = 0.003$, $p < .01$, the main effect of delay, $F(2, 132) = 2.56$, $MSE = 0.10$, $p > .08$, and the interaction between these variables, $F < 1$, failed to reach significance. The Newman–Keuls test showed that participants' reconstruction performance declined across tests in the immediate condition (T1 > T2 = T3) and did not change over tests in the short delay and long delay conditions (T1 = T2 = T3, following both delays). The analysis showed effects of retention interval (top left to bottom right diagonals: immediate > short = long, for each test number) and an effect of repeating testing (column C: .31 > .25 = .24; column B: .31 > .27) that were similar to those described in the picture condition.

The intertest forgetting and recovery data, displayed in Table 4, showed a number of important findings as well. First, as in Experiment 1, participants produced order reminiscence (recovery) between early and late tests in each delay and stimulus condition, and intertest recovery remained constant over these tests and conditions. A Newman–Keuls test revealed that intertest forgetting was lower on late tests than on early tests in the immediate conditions, but intertest forgetting did not change significantly over tests in the short and long delay conditions. These data are consistent with the idea that intertest forgetting declines on late tests because of the negatively accelerated nature of the forgetting function. If this were the case, one would expect to see a decline in forgetting in the immediate condition, but not in the delayed conditions because performance would be closer to the asymptotic segment of the forgetting function.

*Multidimensional order analyses.* As in the first experiment, the reconstruction data were scored with regard to the list and within-list position dimensions. The data from this analysis are displayed in Table 5 as a function of delay condition (immediate vs. delayed), dimension

[4]This positive effect of repeated testing also could be seen as a type of hypermnesia, although not in the traditional sense of the term. Typically, hypermnesia is described as a within-subjects phenomenon (e.g., a participant performs better on Test 3 than on Test 1). The data from Experiment 2 show between-subjects hypermnesia: Performance is better for the group of participants that received two prior tests than for the groups that received either one or no prior tests.

TABLE 5
Serial position functions[a] in Experiment 2

| *Condition* | *Dimension* | *Test* | *Serial position* *1* | *2* | *3* | *4* | *5* | *Mean* |
|---|---|---|---|---|---|---|---|---|
| Immediate | List | 1 | .77 | .47 | .39 | .42 | .45 | .50 |
| | | 2 | .76 | .44 | .36 | .36 | .41 | .46 |
| | | 3 | .76 | .45 | .36 | .35 | .41 | .46 |
| | Change (T3 – T1) | | –.01 | –.02 | –.03 | –.07 | –.04 | –.04 |
| | Within-list | 1 | .69 | .55 | .43 | .46 | .53 | .53 |
| | | 2 | .66 | .54 | .38 | .38 | .51 | .49 |
| | | 3 | .64 | .51 | .41 | .39 | .51 | .49 |
| | Change (T3 – T1) | | –.05 | –.04 | –.02 | –.07 | –.02 | –.04 |
| Delayed | List | 1 | .73 | .43 | .31 | .33 | .39 | .44 |
| | | 2 | .72 | .43 | .31 | .30 | .39 | .43 |
| | | 3 | .71 | .43 | .31 | .29 | .39 | .43 |
| | Change (T3 – T1) | | –.02 | .00 | .00 | –.04 | –.00 | –.01 |
| | Within-list | 1 | .65 | .48 . | .40 | .37 | .44 | .47 |
| | | 2 | .63 | .48 | .40 | .36 | .45 | .46 |
| | | 3 | .63 | .48 | .39 | .36 | .44 | .46 |
| | Change (T3 – T1) | | –.02 | .00 | –.01 | –.01 | .00 | –.01 |

[a]Collapsed across stimulus condition.

(list vs. within-list), serial position (1–5), and test number (1–3).[5] Four separate 3 (test number) × 5 (serial position) ANOVA tests were performed on the mean proportion of correctly reconstructed items for the list and within-list position dimensions of each delay condition.

In the immediate condition, the list dimension ANOVA revealed reliable main effects of test number, $F(2, 178) = 6.04$, $MSE = 0.03$, $p < .01$, and serial position, $F(4, 356) = 63.79$, $MSE = 0.12$, $p < .001$, these variables did not interact, $F < 1$. The list serial position functions were bow-shaped, and overall performance declined across tests by 4%; this drop in performance was distributed uniformly across the lists. Similarly, the within-list position ANOVA indicated reliable main effects of test number, $F(2, 178) = 10.20$, $MSE = 0.03$, $p < .001$, and serial position, $F(4, 356) = 48.71$, $MSE = 0.06$, $p < .001$. The interaction between these variables, however, did approach significance, $F(8, 712) = 1.94$, $MSE = 0.01$, $p < .06$. The serial position functions showed marked primacy and recency effects and net performance dropped by 4% across tests. The locus of the marginally significant interaction is not readily obvious in the data; performance seemed to decline regularly across the serial positions. Also,

[5]As in Experiment 1, the data in each delay condition were collapsed across stimulus condition (pictures vs. words). In addition, the data patterns in the short and long delay conditions did not differ, and, for clarity, these data were combined to form one data set (i.e., delayed condition).

the amounts of intertest forgetting and recovery along the within-list and list dimensions closely resembled those reported in the immediate condition of Table 4.

Turning to the delayed condition, the list dimension ANOVA revealed a significant main effect of serial position, $F(4, 716) = 159.13$, $MSE = 0.09$, $p < .001$, but the main effect of test number, $F < 1.5$, and the Test Number × Serial Position interaction, $F < 1$, were not statistically reliable. The within-list position ANOVA showed a similar pattern of results; the main effect of serial position was reliable, $F(4, 716) = 105.26$, $MSE = 0.05$, $p < .001$, whereas the main effect of test number and the interaction were not, $F$'s $< 1.2$. As in the immediate condition, the serial position functions for the list and within-list position dimensions displayed bow-shaped forms. In contrast, however, overall performance did not change across tests in either dimension. Intertest forgetting and recovery values for the list and within-list dimensions mimicked those displayed in the delayed conditions of Table 4.

*Summary*. The second experiment provided a number of important findings. First, the immediate condition of Experiment 2 replicated the results reported in Experiment 1. When participants received three subsequent reconstruction tests that started immediately after presentation, they showed order reminiscence (recovery) and forgetting between tests, but tended to forget more than they recovered over these tests. Intertest recovery remained constant across early and late tests but forgetting declined on late tests. The multidimensional analyses also replicated the findings of Experiment 1: Overall performance declined along both the list and the within-list position dimensions over tests, and this drop was distributed relatively evenly across the lists and serial positions.

Second, the experiment showed that performance in the short and long delay conditions reached asymptote after about 9 minutes. Participants produced both order recovery and forgetting between tests, but these factors completely balanced one another out—net performance did not change across tests in these conditions. The multidimensional analysis confirmed these findings along the list and within-list position dimensions: There was no change in performance over the three tests in either dimension.

Finally, this examination revealed that both the length of the retention interval and the presence of repeated tests affected reconstruction performance: Participants' performance in this task decreased as the length of the retention interval increased, and repeated testing reduced the net amount of forgetting across tests. These findings differ from those reported by Roediger and Payne (1982). In their study, recall performance depended directly on repeated testing and retrieval practice, but was not affected by retention interval. Although unexplained, these conflicting patterns do represent yet another dissociation between item and order retention: Both retention interval and repeated testing are important determinants of final performance on repeated reconstruction tests, whereas retention interval is not important on repeated recall tests.

## EXPERIMENT 3

In recent years, researchers have become increasingly interested in the role of encoding processes (relational and item-specific processing, specifically) in producing hypermnesia and other repeated testing effects (e.g., Burns, 1993; Klein, Loftus, Kihlstrom, & Aseron, 1989; McDaniel et al., 1998; Otani et al., 1995). McDaniel et al. (1998), for example, demonstrated that relational and item-specific processing have powerful effects on memory performance over repeated recall tests. They showed that relational processing reduced intertest forgetting and produced greater hypermnesia on earlier than on later test trials, whereas item-specific processing enhanced item recovery and produced greater hypermnesia on later tests. From these results, they were able to gain insights into how encoding processes might affect memory performance across repeated tests; they concluded that relational encoding provided participants with a consistent retrieval plan from the outset, while item-specific encoding provided a richer, more recoverable, memory trace. Clearly, then, one must consider both encoding and retrieval processes in order to better understand repeated testing performance.

In the first two experiments, retrieval factors (e.g., type of test, number of tests, delay of tests) were manipulated almost exclusively, with the only encoding variable being the type of stimulus: picture or word. The third experiment further explored the role of encoding processes in repeated order tests. In this experiment, participants were asked to use either a relational or an item-specific processing strategy as they viewed a set of 25 items, which were either pictures or words. Following the processing task, the participants were surprised with three consecutive reconstruction of order tests. Of central interest was how these encoding tasks affected order retention across tests.

## Method

### *Participants and apparatus*

A total of 180 Purdue University undergraduates participated for course credit in an introductory psychology course. Groups of 10 or fewer participants were tested in sessions lasting approximately 60 minutes. The stimuli were presented and controlled by an IBM-compatible computer.

### *Materials and design*

Participants were assigned to one of four experimental conditions (picture/relational, word/relational, picture/item-specific, or word/item-specific) based on their order of arrival to the experiment. In all, there were 45 participants assigned to each of the four conditions. A set of 25 items—5 items from each of five categories—were drawn from Hunt and Hodge (1971), and pictures of these items were selected from Microsoft ClipArt and Blast Software's 10,000 Graphics Pack, Volume 1. The 25 items were arranged into five lists of 5 items. Across participants, two sequences of the items were used. Each sequence was randomly determined with the restriction that no two items from the same category could occur successively in a sequence.

### *Procedure*

An incidental-learning procedure was employed in this experiment because it was necessary to control how participants processed the stimuli. Participants were told that the purpose of the experiment was to investigate characteristics of common words or pictures; they were not informed about the reconstruction tests that followed list presentation. In the relational condition, participants viewed the set of 25 items and were asked to identify the item's appropriate category. They were given a sheet of paper with the names of the five categories listed at the top and a series of 5 blanks next to the numbers 1–5 below. As each item was presented, the participants were asked to write down the number of the category (1–5) that each item belonged to. In the item-specific condition, participants were asked to rate the pleasantness of each item on a scale of 1–5 (1= very unpleasant, 5 = very pleasant) as it was presented.

These ratings were recorded on a sheet of paper similar to that given in the relational condition except that, at the top of the sheet, the pleasantness scale was given instead of the category names.

The stimuli were projected onto a screen at a rate of one item every 5 s, with a 10-s blank interval separating each of the five lists. Immediately following the final item, participants were surprised with their first of three consecutive reconstruction tests. From this point onward, the testing procedure matched that of Experiments 1 and 2.

## Results and discussion

The net reconstruction levels for pictures and words on the three tests following relational and item-specific processing are displayed in Table 6. As in the first two experiments, an item was scored as correct if it was placed into its original absolute position within the set of 25 items (i.e., proper position within proper list). Of primary interest in this experiment was how these processing tasks affected order retention over tests. To address this question, a 2 (processing condition: relational vs. item-specific) × 2 (stimuli: pictures vs. words) × 3 (test number: 1, 2, 3) mixed-factor ANOVA was performed on the mean number of correctly reconstructed items.

The ANOVA revealed reliable main effects of processing condition, $F(1, 176) = 42.43$, $MSE = 0.05$, $p < .001$, stimuli, $F(1, 176) = 21.45$, $MSE = 0.05$, $p < .001$, and test number, $F(2, 352) = 21.32$, $MSE = 0.004$, $p < .001$, the processing condition × stimuli interaction was also reliable, $F(1, 176) = 10.06$, $MSE = 0.004$, $p < .01$. The remaining interactions failed to reach significance, all $F$'s < 1.5. Reconstruction performance was more accurate following item-specific processing than relational processing (e.g., see Hunt & Einstein, 1981, for a similar advantage of item-specific processing with a related list). The data also showed a picture superiority effect—participants reconstructed the order of the pictures better than that of the words. The magnitude of this effect, however, varied with processing condition—the picture superiority effect was larger in the item-specific condition than in the relational condition.

Importantly, net performance declined significantly over tests by 3% and 4%, for pictures and words in each processing condition. Of the 90 participants in the item-specific conditions,

TABLE 6
Net reconstruction levels and rates of intertest recovery and forgetting in Experiment 3

| *Dependent measure* | *Tests* | *Relational* | | *Item-specific* | |
|---|---|---|---|---|---|
| | | *Pictures* | *Words* | *Pictures* | *Words* |
| Net performance | 1 | .21 | .19 | .38 | .24 |
| | 2 | .17 | .14 | .36 | .22 |
| | 3 | .18 | .15 | .35 | .20 |
| | Change (T3 – T1) | –.03 | –.04 | —.03 | –.04 |
| Recovery | Early (T1:T2) | .04 | .04 | .07 | .07 |
| | Late (T2:T3) | .04 | .04 | .06 | .06 |
| Forgetting | Early (T1:T2) | .08 | .09 | .10 | .10 |
| | Late (T2:T3) | .04 | .03 | .06 | .07 |

50 showed a decline in performance, 21 showed the opposite effect, and there were 19 ties. In the relational conditions, overall performance dropped for 48 of the 90 participants, 19 showed the opposite pattern, and there were 23 ties. Participants showed intertest forgetting across early and late tests, and a Newman-Keuls analysis revealed that there was significantly less forgetting on late tests than on early tests in all conditions except for one—the word-item-specific condition. The analysis also revealed that relational and item-specific processing produced comparable amounts of intertest forgetting on early tests, but they produced different amounts of forgetting on late tests; intertest forgetting was lower in the relational condition, which is consistent with McDaniel et al.'s (1998) data that showed a greater reduction in intertest forgetting with relational processing. It is important to note that the aforementioned finding was not due to a floor effect in the relational condition, as performance was well above chance.

Participants also produced order reminiscence (recovery) between early and late tests in each processing and stimulus condition. The Newman–Keuls analysis revealed that, for a given processing condition, intertest recovery remained constant over the three tests. However, the analysis also revealed that item-specific processing produced significantly more recovery than did relational processing on early tests; this difference was only marginally significant on late tests (both $p$s $< .10$). McDaniel et al. (1998; Exp. 3) showed a similar result in recall and suggested that a memory trace is more recoverable following item-specific processing because this processing produces a memory trace with a richer set of encoded attributes; an item with many attributes has a retrieval advantage over an item with fewer attributes because a given set of retrieval cues is more likely to share some attributes with a richer trace, thus increasing the likelihood of retrieving that item. It seems that item-specific processing might lead to a recovery advantage in reconstruction for this same reason. Even though the list items are re-presented at test in reconstruction, one must still access the appropriate representations in memory; after all, the items themselves serve only as cues and do not guarantee access to the proper mnemonic representations (Neath, 1997; Tulving, 1983). Accordingly, there should be a greater likelihood of accessing the original processing record, and hence a greater probability of recovery, following the richer, item-specific encoding than following the more deficient, relational encoding.

*Multidimensional order analyses.* As in the first two experiments, the reconstruction data were scored with regard to the list and within-list position dimensions. The data from this analysis are displayed in Table 7 as a function of processing condition (relational vs. item specific), dimension (list vs. within-list), serial position (1–5), and test number (1–3).[6] Four separate 3 (test number) × 5 (serial position) ANOVA tests were performed on the mean proportion of correctly reconstructed items for the list and within-list position dimensions of each processing condition.

In the relational processing condition, the list dimension ANOVA revealed reliable main effects of test number, $F(2, 178) = 3.27$, $MSE = 0.03$, $p < .05$, and serial position, $F(4, 356) =$

[6]As in the first two experiments, the data in each processing condition were collapsed across stimulus condition (pictures vs. words); there were no detectable differences in the list and within-list position data patterns for pictures and words.

TABLE 7
Serial position functions[a] in Experiment 3

| Condition | Dimension | Test | Serial position 1 | 2 | 3 | 4 | 5 | Mean |
|---|---|---|---|---|---|---|---|---|
| Relational | List | 1 | .46 | .30 | .31 | .33 | .45 | .37 |
| | | 2 | .45 | .30 | .26 | .32 | .41 | .35 |
| | | 3 | .45 | .28 | .28 | .30 | .41 | .34 |
| | Change (T3 – T1) | | –.01 | –.02 | –.03 | –.03 | –.04 | –.03 |
| | Within-list | 1 | .41 | .32 | .32 | .31 | .36 | .34 |
| | | 2 | .40 | .30 | .29 | .32 | .33 | .33 |
| | | 3 | .40 | .30 | .31 | .30 | .33 | .33 |
| | Change (T3 – T1) | | –.01 | –.02 | –.01 | –.01 | –.03 | –.01 |
| Item-specific | List | 1 | .70 | .46 | .33 | .36 | .60 | .49 |
| | | 2 | .66 | .42 | .35 | .35 | .55 | .47 |
| | | 3 | .65 | .42 | .31 | .33 | .55 | .45 |
| | Change (T3 – T1) | | –.05 | –.04 | .02 | –.03 | –.05 | –.04 |
| | Within-list | 1 | .65 | .49 | .45 | .40 | .56 | .51 |
| | | 2 | .61 | .43 | .42 | .38 | .54 | .47 |
| | | 3 | .60 | .42 | .41 | .37 | .52 | .47 |
| | Change (T3 – T1) | | –.05 | –.07 | –.04 | –.03 | –.04 | –.04 |

[a]Collapsed across stimulus conditions.

26.15, *MSE* = 0.07, $p < .001$; these variables did not interact, $F < 1$. The list serial position functions were bow-shaped, and overall performance declined across tests by 3%; this drop in performance was distributed uniformly across the lists. The within-list ANOVA indicated a reliable main effect of serial position, $F(4, 356) = 8.31$, *MSE* = 0.06, $p < .001$, but neither the main effect of test number nor the Test Number × Serial Position interaction reached significance, *F*'s < 1. The serial position functions showed clear primacy and recency effects, and net performance did not change across tests. This is the first time that within-list performance did not decline across tests when participants received their first test immediately after presentation; the locus of these disparate results is not readily obvious.

In the item-specific condition, the list dimension ANOVA revealed reliable main effects of test number, $F(2, 178) = 5.36$, *MSE* = 0.03, $p < .01$, and serial position, $F(4, 356) = 81.97$, *MSE* = 0.07, $p < .001$, and these variables did not interact, $F < 1$. The within-list position ANOVA showed a similar pattern of results; the main effects of test number, $F(2, 178) = 6.46$, *MSE* = 0.03, $p < .01$, and serial position, $F(4, 356) = 52.39$, *MSE* = 0.05, $p < .001$, were reliable, $F(4, 716) = 105.26$, *MSE* = 0.05, $p < .001$, and the interaction of these variables was not, $F < 1$. As in the relational condition, the serial position functions for the list and within-list position dimensions displayed bow-shaped forms. In contrast, however, overall performance declined significantly across tests in both dimensions; this drop was spread evenly across lists and within-list positions. For all conditions, the amounts of intertest forgetting and recovery closely resembled the values reported in the corresponding conditions of Table 6.

*Summary.* The third experiment examined the role of encoding processes in repeated order tests and provided a number of interesting results. First, item-specific processing supported a higher level of reconstruction performance than did relational processing. Although both processing conditions showed a similar drop in performance over tests, they displayed different patterns of order loss and recovery between tests. Most notably, item-specific encoding produced more intertest recovery than did relational encoding across early and late tests. Item-specific processing also yielded more forgetting between late tests than did relational processing. These effects are consistent with McDaniel et al.'s (1998) suggestions that item-specific encoding enhances intertest recovery, and relational processing reduces intertest forgetting. The multidimensional analyses showed that overall performance declined along the list and within-list position dimensions following item-specific encoding and only along the list dimension following relational encoding.

## GENERAL DISCUSSION

The present experiments established that repeated testing affects item and order retention differently: Participants produced hypermnesia with repeated free recall tests whereas participants' net performance declined across successive order reconstruction tests. These experiments demonstrated that order performance declined over tests with a variety of encoding conditions (pictures, words, and relational and item-specific processing) and retrieval conditions (intentional and incidental learning). Although overall performance dropped across tests, participants showed reliable order recovery (reminiscence) between tests—they were able to remember previously forgotten order information. Thus, it appears that reminiscence is a general property of repeated testing, having been found with free recall, cued recall, recognition, and reconstruction, whereas hypermnesia is a more restricted phenomenon, having only been found with free and cued recall. Intertest recovery remained constant across early and late tests (about 4%) while intertest forgetting typically declined on later tests. The multidimensional order analyses revealed that reconstruction performance generally declined over tests along both the list and the within-list position dimensions as well, and these drops in performance were generally distributed evenly across lists and within-list positions.

Performance appeared to decline across tests primarily because of the increased retention interval associated with repeated testing. Experiment 2 showed that reconstruction performance declined as the length of the retention interval increased. Repeated testing itself did not adversely affect performance, quite the contrary in fact. Performance following a constant retention interval was reliably better when participants received prior tests. It seems, then, that the repeated tests themselves were not necessarily responsible for the loss—the increased retention interval was to blame.

In addition to these findings, the present experiments revealed a number of other important and novel findings. First, participants demonstrated robust picture superiority effects in reconstruction: Pictures were reordered better than words in each experiment. Second, participants' reconstruction performance was better following item-specific processing than following relational processing. Moreover, item-specific processing produced significantly more recovery (reminiscence) between tests and significantly more forgetting on late tests than did relational processing.

## Theoretical interpretations

Currently, the three most prominent accounts of reminiscence and hypermnesia, namely the cumulative recall hypothesis (Roediger, 1982; Roediger, Payne, Gillespie, & Lean, 1982), the retrieval dynamics account (Payne et al., 1993), and the relational/item-specific account (McDaniel et al., 1998), do not explicitly address the retention and recovery of order information and, therefore, are silent on the effects of repeated testing on memory for order. In order to be a parsimonious explanation of reminiscence and hypermnesia, these theories will need to be modified to address a number of findings. First, they will need to account for the decline in net order performance across tests, while allowing for small, but consistent, recovery between tests and a slightly greater amount of forgetting that declines on late tests. Given that some degree of intertest forgetting is at work in each of these theories, one could account for these findings by assuming that order information is lost more rapidly than item information, such that the amount of forgetting between tests will always exceed the amount of recovery. Although this technique appears easy and straightforward, it is not a trivial modification. Indeed, such a change would require clear explanation and delineation of the conditions in which the intertest forgetting rate will be greater than the intertest recovery rate, and vice versa, so that a priori predictions could be generated.

In addition, these theories will need to include processes or mechanisms that allow performance to change over the course of the retention interval (e.g., through changing context cues) and across tests (e.g., through retrieval practice), as well as reproduce how different encoding variables (e.g., relational and item-specific processing) affect order retention across tests. Each of these theories will require considerable modification to address the aforementioned findings with the exception of the relational/items-specific account (McDaniel et al., 1998), which might be able to account for the encoding effects with only minor modifications. McDaniel et al. showed that, in a free recall task, relational encoding reduced intertest forgetting, and item-specific encoding enhanced item recovery. These results seem consistent with the findings reported in Experiment 3—less forgetting between tests following relational encoding and greater recovery following item-specific encoding. To explain their results, McDaniel et al. suggested that relational encoding provides participants with a consistent retrieval plan, which reduces forgetting, and that item-specific encoding produces a richer, more recoverable, trace, which boosts recovery. If one assumes that relational and item-specific processing do not differ in the degree of order processing that they encourage, then their explanation seems reasonable when applied to the realm of repeated order testing.

It is possible, however, that relational and item-specific processing do in fact differ in the degree of order processing that they encourage. Item-specific processing might encourage participants to focus more on the individual items and less on the ordered relationships among those items (e.g., DeLosh & McDaniel, 1996; McDaniel, DeLosh, & Merritt, 2000; Nairne, Riegler, & Serra, 1991). If this were the case, one might expect item-specific encoding to produce high intertest forgetting (no consistent retrieval plan) and low order recovery (deficient encoding of order information). To the extent that relational processing does not adversely affect order encoding, one might expect lower intertest forgetting (consistent retrieval plan) and higher order recovery (order information encoded more completely). Two potential problems stem from this interpretation. First, any account that predicts better order retention for relational processing than for item-specific processing is not consistent with the results from Experiment 3, which show superior performance following item-specific processing. Second, participants in this situation might focus solely on interrelations among the list items (e.g., notice that the items fall into three different categories) at the exclusion of order information. Clearly, such issues would need to be addressed before McDaniel et al.'s (1998) relational/item-specific approach could provide an adequate account of the current data.

General theories of order memory (e.g., Estes', 1997, perturbation theory; Henson's, 1998, start–end model) are unable to provide a complete account of these effects as well, although they do contain mechanisms that produce reminiscence while allowing net performance to decline across tests. For example, in perturbation theory, performance declines over tests because there is more time, and hence more opportunity, for items to perturb out of their proper positions. Reminiscence occurs when an item that was recalled out of its proper position at Test 1 perturbs back into its appropriate position at Test 2. In the start–end model, performance declines because, when an item is retrieved in the model, the position of that item's token in memory is recoded to its output position (e.g., if Token A was originally in Position 1 and is subsequently recalled in Position 2, then the position of Token A is recoded to correspond with its output position—2). Thus, once a mistake is made, that mistake is likely to be repeated on later tests. When this process is coupled with the noisy item-selection process, performance typically declines across tests. In addition, performance declines over tests due to the influence of a contextual cue that loses strength as time passes. Reminiscence occurs at random in this model, through the noisy selection process.

Although these models can account for the presence of order reminiscence without hypermnesia, they do not currently contain explicit mechanisms that allow one to manipulate variables such as pictures versus words or item-specific versus relational encoding, nor are these models able to account for the effects of increased retention interval and repeated testing reported in Experiment 2. For example, in the start-end model, performance declines as a function of both the increased retention interval associated with repeated testing and the repeated tests themselves. Although the negative effect of retention interval is consistent with the current findings, the negative effect of repeated testing is not; Experiment 2 showed that repeated testing had a positive effect on order retention by attenuating the net amount of forgetting across tests. The perturbation model is also limited in that it can only account for the main effect of increased retention interval reported in Experiment 2 and not the main effect of repeated testing. In its current form, performance declines over tests because there is more time, and hence more opportunity, for items to perturb out of their proper positions; the repeated tests themselves do not affect performance. Thus, it appears that these models will require revision of some of their fundamental assumptions to account for the full range of data reported in this article.

## Is reconstruction like recognition and/or recall?

With the data from the present experiments, one can also consider whether the processes involved in a reconstruction task tap processes similar to those employed in recognition and/or recall. In an attempt to determine whether the processes involved in reconstruction are similar to those employed in recognition and/or in recall, Whiteman, Nairne, and Serra (1994) examined how word frequency affected participants' performance in recall, recognition, and reconstruction. Previous investigations showed different word frequency effects in recall and recognition: High-frequency words enjoyed an advantage in recall (in pure frequency lists), while low-frequency words were better recognized (in pure or mixed frequency lists, e.g., Gillund & Shiffrin, 1984; Gregg, 1976). They reasoned that an advantage for high-frequency items in a reconstruction task would indicate that recall-like processes are involved in reconstruction, whereas a low-frequency advantage would implicate recognition-like processes. They found that word frequency did not affect reconstruction performance (high- and low-frequency items were reordered equally well) and concluded that the processes involved in reconstruction are different from those in recall and recognition.

Following the logic of Whiteman et al. (1994), the current data seem to suggest that reconstruction might involve some recognition-like processes. As noted earlier, repeated testing produces quite different effects on recall and recognition. While reminiscence is common to both tests, one typically finds hypermnesia with recall and either a decline or no change in performance with recognition. Clearly, the drop in reconstruction performance across tests more closely resembles the pattern of data found in recognition than that found in recall.

One could argue that, because both recognition and reconstruction provide target items as cues at test, they will induce similar retrieval processes and/or strategies. For example, Payne et al. (1993) suggested that, when targets are given as cues, there is little, if any, fluctuation in the cues at retrieval and, consequently, little recovery of new information; a low recovery rate, accordingly, reduces the likelihood that performance will improve across tests. If reconstruction employed a recall-like process, one might have expected greater recovery and an overall improvement in performance over tests. Thus, at least under repeated testing conditions, it seems that similar processes guide performance in both reconstruction and recognition.

## REFERENCES

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

Bjork, E. L., & Healy, A. F. (1974). Short-term order and item retention. *Journal of Verbal Learning and Verbal Behavior*, *13*, 80–97.

Burns, D. J. (1993). Item gains and losses during hypermnesic recall: Implications for the item-specific-relational information distinction. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *19*, 163–173.

Crowder, R. G. (1976). *Principles of learning and memory*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

DeLosh, E. L., & McDaniel, M. A. (1996). The role of order information in free recall: Application to the word-frequency effect. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *22*, 1136–1146.

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

Estes, W. K. (1997). Processes of memory loss, recovery, and distortion. *Psychological Review*, *104*, 148–169.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, *91*, 1–67.

Gregg, V. (1976). Word frequency, recognition, and recall. In J. Brown (Ed.), *Recall and recognition* (pp. 183–216). London: Wiley.

Healy, A. F. (1974). Separating item from order information in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *13*, 644–655.

Henson, R. N. A. (1998). Short-term memory for serial order: The start–end model. *Cognitive Psychology*, *36*, 73–137.

Hunt, K. P., & Hodge, M. H. (1971). Category-item frequency and category-name meaningfulness (m'): Taxonomic norms for 84 categories. *Psychonomic Monograph Supplements*, *4*, 97–121.

Hunt, R. R., & Einstein, G. O. (1981). Relational and item-specific information in memory. *Journal of Verbal Learning and Verbal Behavior*, *20*, 497–514.

Klein, S. B., Loftus, J., Kihlstrom, J. F., & Aseron, R. (1989). Effects of item-specific and relational information on hypermnesic recall. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *15*, 1192–1197.

McDaniel, M. A., DeLosh, E. L., & Merritt, P. S. (2000). Order information and retrieval distinctiveness: Recall of common versus bizarre material. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *26*, 1045–1056.

McDaniel, M. A., Moore, B. A., & Whiteman, H. L. (1998). Dynamic changes in hypermnesia across early and late tests: A relational/item-specific account. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *24*, 173–185.

Murdock, B. B., Jr. (1976). Item and order information in short-term serial memory. *Journal of Experimental Psychology: General*, *105*, 191–216.

Murdock, B. B., Jr., & vom Saal, W. (1967). Transpositions in short-term memory. *Journal of Experimental Psychology*, *74*, 137–143.

Nairne, J. S. (1990). Similarity and long-term memory for order. *Journal of Memory and Language*, *15*, 859–867.

Nairne, J. S. (1991). Positional uncertainty in long-term memory. *Memory & Cognition*, *19*, 332–340.

Nairne, J. S., & Kelley, M. R. (1999). Reversing the phonological similarity effect. *Memory & Cognition*, *27*, 45–53.

Nairne, J. S., Riegler, G. L., & Serra, M. (1991). Dissociative effects of generation on item and order retention. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *17*, 702–709.

Neath, I. (1997). Modality, concreteness, and set-size effects in a free reconstruction of order task. *Memory & Cognition*, *25*, 256–263.

Otani, H., & Hodge, M. H. (1991). Does hypermnesia occur in recognition and cued recall? *American Journal of Psychology*, *104*, 101–116.

Otani, H., & Stimson, M. J. (1994). A further attempt to demonstrate hypermnesia in recognition. *The Psychological Record*, *44*, 25–34.

Otani, H., & St. Louis, P. (1995). A further examination of relational and item-specific processing effects on cued recall hypermnesia. *Journal of General Psychology*, *122*, 355–364.

Otani, H., & Whiteman, H. L. (1994). Cued recall hypermnesia is not an artifact of response bias. *American Journal of Psychology*, *107*, 401–421.

Paivio, A. (1974). Spacing of repetitions in the incidental and intentional free recall of pictures and words. *Journal of Verbal Learning and Verbal Behavior*, *13*, 497–511.

Paivio, A., & Csapo, K. (1974). Picture superiority in free recall: Imagery or dual coding? *Cognitive Psychology*, *5*, 176–206.

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

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

Payne, D. G., Hembrooke, H. A., & Anastasi, J. S. (1993). Hypermnesia in free and cued recall. *Memory & Cognition*, *21*, 48–62.

Payne, D. G., & Roediger, H. L., III. (1987). Hypermnesia occurs in recall but not in recognition. *American Journal of Psychology*, *100*, 145–165.

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

Roediger, H. L., III., & Challis, B. H. (1989). Hypermnesia: Improvements in recall with repeated testing. In C. Izawa (Ed.), *Current issues in cognitive processes: The Tulane Floweree Symposium on cognition*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

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

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

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 174–215.

Tulving, E. (1983). *Elements of episodic memory*. New York, NY: Oxford University Press.

Whiteman, H. L., Nairne, J. S., & Serra, M. (1994). Recognition and recall–like processes in the long-term reconstruction of order. *Memory*, *2*, 275–294.

*Original manuscript received 16 August 2001*
*Accepted revision received 4 March 2002*

# APPENDIX
## Sample reconstruction test

1) _________ _________ _________ _________ _________
2) _________ _________ _________ _________ _________
3) _________ _________ _________ _________ _________
4) _________ _________ _________ _________ _________
5) _________ _________ _________ _________ _________

---

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| bear | hammer | corn | scissors | apple | moon | church | foot | |
| clock | telephone | arrow | button | lamp | door | airplane | chair | |
| violin | bird | truck | kite | spoon | fence | butterfly | ring | pencil |