

## Recognition and Recall-like Processes in the Long-term Reconstruction of Order

Howard L. Whiteman, James S. Nairne, and Matt Serra  
*Purdue University, Indiana, USA*

Four experiments examined whether the recovery of an item's position in a sequence taps processes similar to recognition and/or recall. Across the experiments, subjects either recalled, recognised, or made position judgements about list items that differed in word frequency. Typical word frequency effects were found in recall and recognition, but frequency failed to affect measures of position memory consistently across the four experiments. Despite the apparent procedural similarities across tasks, it appears that the recovery of position information may tap mnemonic processes that are different from those tapped by recognition and recall. Implications of these findings for current models of position memory are discussed.

### INTRODUCTION

Remembering order, or the position of items in a sequence, has been of concern to psychologists for some time. Over the years, order memory has been assessed in a variety of ways, including serial recall, serial anticipation learning, and through measurement of the amount of input-output correspondence in free recall (see Berch, 1979, for a review). Unfortunately, many of these measures have confounded memory for the item's relative position in the sequence with memory for the item itself. For example, in the measures just listed subjects are required to remember list items as well as their order of occurrence. As a result, it can be difficult to untangle the relative contributions of item and order retention in final recall performance.

To eliminate the confounding, researchers have relied primarily on one of three types of tasks. In one task, subjects are required to recall the order of a fixed set of items that does not vary from trial to trial (e.g. Healy, 1974; Lee & Estes, 1977). Because the same items are used on every trial, like the digits 0-9, it is assumed that the recovery of item information is never a concern. In a

---

Requests for reprints should be sent to James S. Nairne, Department of Psychological Sciences, Purdue University, 1364 Psychology Building, West Lafayette, IN 47907-1364, USA.

The authors thank Susan Devine, Amy Gibson, and Kelly Pierce for their assistance in running the experiments. Matt Serra is now at Duke University.

second kind of task, subjects are given the items (which may vary from trial to trial) individually at test, and are required to make position judgements about each (e.g. Schulz, 1955); in this case, item information is provided explicitly at test, so the task requires only the remembering of position. A third technique provides the subject with all of the item information from a trial—usually the presented items in a new random order—and requires reconstruction of the original presentation sequence (e.g. Healy, Fendrich, Cunningham, & Till, 1987; Nairne, 1990, 1991, 1992; Nairne, Riegler, & Serra, 1991; Serra & Nairne, 1993). Again, because the critical items are available at the time of testing, the measure is typically seen as a pure test of position or order memory.

Studies using these tasks have revealed some important facts about how position information is forgotten over time. Specifically, position errors tend to cluster around true serial positions and decline gradually as distance from the true position increases. Several quantitative models have been proposed which can account for these error functions (or positional uncertainty distributions). In the Estes perturbation model (e.g. Lee & Estes, 1977), for example, items are associated to specific positional values during encoding. As time passes, however, with a certain probability each individual position memory perturbs along the position dimension and comes to take on the value of a neighbouring position. Perturbation theory accounts for the shape of the error distributions, as well as for the general bow-shaped serial position function that is the hallmark of immediate order retention (Estes, 1972; Lee & Estes, 1977, 1981).

There is a fair amount of empirical support for such quantitative models, and the perturbation model in particular (e.g. Healy et al., 1987; Jahnke, Davis, & Bower, 1989; Nairne, 1991, 1992; Nairne & Dutta, 1992), but significant questions remain about the tasks that are typically used to examine the characteristics of order retention. For example, hardly any studies have considered the empirical relationship among tasks that measure position memory and more traditional memory measures such as recall and recognition. Would we expect the recovery of an item's position in sequence to be affected by the same experimental manipulations that affect recall and/or recognition? The answer to such a question is important because well-developed models of order memory, like perturbation theory, might potentially be extended to handle more general phenomena like recall or recognition; or perhaps existing models of recognition and recall (e.g. Gillund & Shiffrin, 1984; Hintzman, 1988) could be easily reformulated to handle order phenomena.

One way to determine whether position judgements tap mnemonic processes similar to those underlying recall and recognition is to manipulate the relative frequency of the to-be-remembered list items. High frequency words are typically recalled better than low frequency words (in pure frequency lists), whereas low frequency words are better recognised (in either pure or mixed frequency lists; see, for example, Gregg, 1976; Gillund & Shiffrin, 1984). In a task like reconstruction, which provides a relatively pure index of position

memory, an advantage for high frequency items would therefore implicate recall-like processes in the position judgement. If the opposite finding was obtained, it would suggest that recognition-like processes might be involved. It is worth noting that this same manipulation and reasoning has recently been employed to determine the processes involved in associative and cued recognition tasks (Clark, 1992; Clark, Hori, & Callan, 1993; Clark & Shiffrin, 1992; Hockley, 1991).

One can easily make the argument that memory for position is based on a recall-like process. Subjects might recall the list items, in order, and determine position by noting the target item's location in the recalled sequence. It is commonly believed that lists of high frequency words are better recalled because inter-item links are easier to form among high frequency words (Deese, 1960; Sumbly, 1963). If reconstruction requires the subject to recall the items internally, in order to determine relative position, then the task might be expected to show an advantage of high over low frequency words. In support of this view, Watkins and Watkins (1977) found a high frequency advantage at early list positions during serial recall, but no frequency difference in the recency portions of the list (see also Gregg, Montgomery, & Castaño, 1980). However, it is unclear what to conclude from these experiments because, as mentioned earlier, item and order information are confounded in serial recall tasks. It is difficult to know whether the high frequency advantage is truly due to order memory, or simply due to an enhanced ability to remember the items themselves.

On the other hand, at least superficially, the procedures of reconstruction closely resemble the task requirements of a recognition test. Subjects are given the items at test, and the task is to make a judgement about occurrence. The only difference is that traditional recognition tasks require coarse-grained judgements (did the item occur in the experiment?) and reconstruction tasks require fine-grained judgements (in which position in the list did the item occur?). Indeed, Crowder (1976, p. 371) has suggested that "reconstruction, for example, can be described as a form of recognition test in which memory for order is examined". This kind of reasoning predicts that low frequency words should show an advantage in a task like reconstruction.

In this article, we report the results of four experiments examining the effect of word frequency on reconstruction, recall, and recognition tasks. In Experiment 1, subjects were given lists containing either high or low frequency words, and the lists were followed by a standard reconstruction task. In Experiment 2, recall and reconstruction were compared directly. In Experiment 3, subjects received lists followed by single words—in one block of trials, subjects were asked to make recognition judgements about the words; in another block, the task was to retrieve the item's correct serial position. Finally, in Experiment 4, subjects were once again required to make judgements about occurrence, but at different levels of specificity—the decision was either coarse-

grained as in recognition (did the item occur in the list?); fine-grained, as in reconstruction (name the serial position of the item); or medium-grained (did the item occur in the first or second half of the list?). In all cases, the question of interest centred on the effect of word frequency on final task performance.

## EXPERIMENT 1

In Experiment 1, subjects were presented with lists of eight items, composed of either high or low frequency words. The design was within-subject with respect to frequency—all subjects received lists containing high and low frequency words—but a given list was restricted to one type of frequency. The final item of each list was immediately followed by 12 seconds of a digit tracking distractor task to clear any residual effects of short term memory. After the distraction interval, the items from the list were presented again, but in a new random ordering. The subject's task was to reconstruct the order of the original presentation. If reconstruction is based on a recall-like process, high frequency lists should show better overall performance. Alternatively, a low frequency advantage should be found if reconstruction is more like recognition than recall.

## Method

*Subjects and Apparatus.* The subjects were 32 introductory psychology students who participated as part of a course requirement. Everyone was tested individually in a single session lasting approximately one hour. Stimulus materials and instructions were presented and controlled by an IBM-compatible computer.

*Materials.* A total of 384 words were chosen from the Toglia and Battig (1978) semantic word norms to serve as an item pool for all four experiments. These words were mono- or bisyllabic, of moderate to high imagery (range = 3.88 to 6.61), and ranged from four to eight letters in length. One half of these items were of high frequency (occurrence > 60 per million; mean = 158.99); the other half were rated low in frequency (occurrence < or = 10 per million; mean = 3.99), based on the Francis and Kucera (1982) frequency norms. Imagery values were similar across the two item pools (high frequency mean = 5.57; low frequency mean = 5.40).

From this item pool, 14 high and 14 low frequency lists, of eight items each, were created. Items were assigned in a pseudorandom fashion to both list and within-list positions. Two restrictions were placed on the item to list assignments: first, an attempt was made to minimise semantic or orthographic overlap within a list. Second, no list could be highly deviant in terms of mean frequency or imagery values compared to other lists of the same frequency type. Any violation of either of these restrictions was corrected by returning to the sampling pool and replacing any deviant item with a new, randomly chosen

item. A fixed presentation order for each of the high and low frequency lists was then randomly determined. However the order of occurrence of high and low frequency lists was randomly determined for each subject. Thus, while the order of the lists *within* each frequency was the same for all subjects, the order of lists *across* frequencies could differ.

Thirty two words of medium frequency were chosen for four practice trials. These trials were not scored.

*Procedure.* Following practice, subjects received 28 eight-item lists. Each trial began with the word *ready*, presented for two seconds along with a tone. Next, the eight list items appeared one at a time, for two seconds each, in the centre of the computer screen. Subjects were asked to read each word aloud as it appeared. Immediately following the final item in each list, single digits, randomly chosen from the set 1–9, began to appear in the centre of the screen for 500ms each. Subjects were required to read each of these digits aloud. This distractor task lasted 12 seconds and was followed by a reconstruction test.

For the reconstruction task, the eight words from the trial were presented in a new random order and appeared on a single line centred in the top half of the terminal screen. Below each item, a corresponding letter (a–h) appeared. In the centre of the terminal screen, there appeared a row of eight blank lines and a flashing cursor. Subjects were told to reconstruct the original order of presentation by placing the letter corresponding to a given word in its appropriate blank line. Subjects were allowed to fill the blanks in any order they wished. Cursor location was subject-controlled via the left and right arrow keys on the computer keyboard. On locating a desired position, a letter was placed by pressing the ENTER key and then the desired letter on the keyboard. Subjects were allowed to work at their own pace. Following the eighth response on each test, instructions appeared asking the subject to press the space bar when ready to begin the next trial.

## Results and Discussion

Only those items that were placed in their correct within-list positions were counted as correct. Statistical reliability was measured at the  $P < 0.05$  level for all analyses. The proportion of correct responses as a function of word frequency and serial position is shown in Fig. 1. A 2 (frequency)  $\times$  8 (serial position) analysis of variance (ANOVA) revealed a reliable effect of serial position,  $F(7,217) = 73.05$  ( $MSE = 0.021$ ). Neither the main effect of frequency,  $F(1, 31) = 1.72$  ( $MSE = 0.034$ ), nor the frequency by serial position interaction,  $F(7,217) = 1.59$  ( $MSE = 0.014$ ), approached significance. Proportion correct for the lists containing high frequency words (0.51) closely matched performance for the low frequency lists (0.49). Of the 32 subjects in this experiment, 17 showed a high frequency advantage, 13 showed a low frequency advantage, and there were two ties.

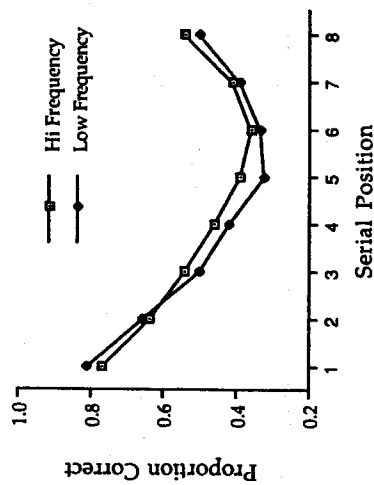


FIG. 1. Proportion correct reconstruction as a function of word frequency and serial position.

The individual position error gradients were also examined (due to a programming error, the data of two subjects were unavailable for determining the response error gradients). These error gradients were of typical form—regardless of frequency, items were most likely to be placed in their correct within-list position; but, when an item was placed incorrectly, it was more likely to be placed in a position adjacent to its correct list position.

The results of this experiment suggest that variations in word frequency have no effect on a reconstruction task, which is designed to measure pure memory for order. The overall serial position curves were nearly identical for high and low frequency words, and there were no differences in the distribution of position errors across the two frequency levels. There are, however, two alternative explanations for the absence of frequency effects in this experiment. First, the range of frequencies that we employed might not be sufficient for producing word frequency effects in the more traditional memory tasks. Second, it is possible that our inability to find frequency effects was due to our use of relatively short lists; most experiments investigating frequency effects have used long lists, usually with more than 10 items (although see Gregg et al., 1980; Watkins & Watkins, 1977). Experiment 2 sheds light on the feasibility of these explanations by having subjects perform both free recall and reconstruction trials.

## EXPERIMENT 2

Experiment 2 followed the procedures of Experiment 1 in all respects except for testing. Subjects in Experiment 2, on a random half of the trials, were required to free recall the critical words instead of reconstructing the presentation order. Based on the findings of Experiment 1, we predicted no effect of word frequency on the reconstruction trials, but an advantage for the high frequency words during free recall.

## Method

**Subjects and Apparatus.** The subjects were 42 undergraduates who participated for course credit. Each subject was tested individually in a single session lasting approximately one hour. Stimulus materials and instructions were presented and controlled by an IBM-compatible computer.

**Materials.** Thirty-two eight-item lists, half with high frequency words and half with low frequency words, were used in the experiment. List construction followed the same restrictions described in Experiment 1. The presentation order of the high and low frequency lists was randomly determined by computer for each session. Each list was followed, one half of the time, by either a free recall or reconstruction test. To ensure that free recall and reconstruction tests were distributed evenly throughout the session, the lists of each frequency were randomly assigned, in blocks of two, to one of the two test types. These list-test combinations were counterbalanced across subjects such that each list followed an equal number of times by both test types. Prior to starting the actual experimental trials, subjects performed two practice trials of each test type. These trials were not scored.

**Procedure.** The reconstruction trials mimicked the procedures of Experiment 1 in all respects. Subjects were presented with the eight words, individually for two seconds apiece, and the reconstruction test was preceded by 12 seconds of a digit-tracking distractor task. On the free recall trials, the final distraction digit was immediately replaced by the message "RECALL NOW". Subjects were asked to write, in any order, all of the words that they could remember from that trial. A booklet of blank half-sheets of paper was provided for the written recall. The duration of each test interval was subject-controlled. After completing their recall, subjects were instructed to turn to a new page in their booklet and to press the space bar when ready to begin the next trial.

## Results and Discussion

Items were scored as correct in the reconstruction task only if they were placed in their correct within-list position. In free recall, order of recall was not a criterion for accuracy. However, items recalled from previously tested lists were not counted as correct (the number of these intrusions was small, and similar across both frequency conditions).

**Free Recall.** The free recall results are displayed in the top panel of Fig. 2, which gives proportion correct as a function of word frequency and serial position. An ANOVA revealed reliable effects of frequency,  $F(1, 40) = 12.46$  ( $MSE = 0.050$ ) and serial position,  $F(7, 280) = 10.47$  ( $MSE = 0.033$ ); the

frequency by serial position interaction did not reach conventional levels of significance,  $F(7,280) = 1.84$  ( $MSE = 0.023$ ). There was an additional three-way interaction involving frequency, position, and the counterbalancing of lists, but careful inspection of the data produced no obvious or meaningful interpretation.

The proportion of high frequency words recalled (0.51) exceeded that of low frequency words (0.45). Of the 42 subjects, 29 showed a high frequency recall advantage, 12 showed the reverse pattern, and there was one tie. Thus, the typical word frequency effect was produced in free recall, which confirms the effectiveness of our frequency manipulation and presentation procedures.

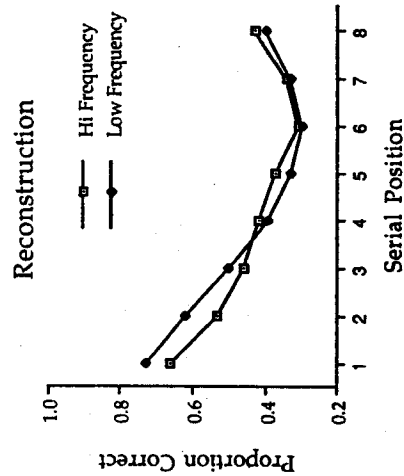
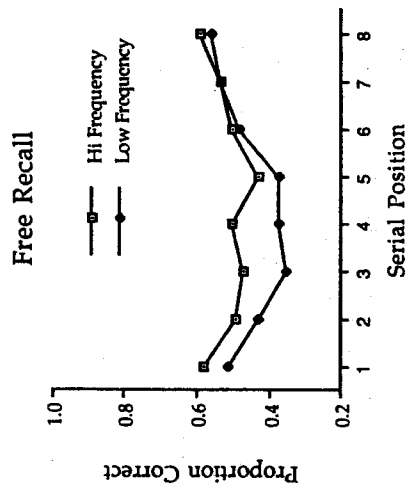


FIG. 2. Proportion correct recall and reconstruction as a function of word frequency and serial position.

**Reconstruction.** The proportions of correct responses for high and low frequency words at each serial position are displayed in the lower panel of Fig. 2. The ANOVA on these data revealed a reliable effect of serial position,  $F(7,280) = 52.83$  ( $MSE = 0.028$ ), but no significant effect of word frequency,  $F(1, 40) < 1$ . The frequency by serial position interaction was also significant,  $F(7,280) = 3.02$  ( $MSE = 0.017$ ) as was the three-way interaction of frequency, counterbalancing, and serial position,  $F(7,280) = 2.74$  ( $MSE = 0.017$ ). Once again, close examination of the three-way interaction produced no meaningful interpretation.

The serial position curves for reconstruction show a primary advantage for low frequency words. This pattern was evident across the counterbalancing of lists, and accounts, in part, for the two-way interaction of frequency and serial position. Overall, however, performance on the high frequency lists (0.44) closely matched performance on low frequency lists (0.45). For the individual data, 20 of the 42 subjects showed a high frequency advantage, 20 showed a low frequency advantage, and there were two ties. Thus, even though the manipulation of frequency was sufficient to produce highly reliable effects in free recall, no consistent effect of word frequency was produced in reconstruction. The results of both Experiments 1 and 2 suggest, then, that processes underlying the recovery of position may differ from those involved in recall. In our next experiment, we sought to compare position judgements directly with recognition.

### EXPERIMENT 3

As in the earlier experiments, subjects in Experiment 3 received a number of eight-item lists followed by a 12-second digit-tracking distractor task. At test, however, subjects received separate blocks of recognition and position judgement test trials. In both tasks, the distractor period ended with the presentation of single words from the list. For the position block, subjects made serial position judgements about each word individually (which differed from the reconstruction procedure used in the previous experiments); for the recognition block, items from the list were mixed with word distractors and the task required simple old/new recognition judgements.

### Method

**Subjects and Apparatus.** Participants were 44 students from the introductory psychology subject pool. Everyone was tested individually in a single session lasting approximately one hour. Stimulus materials and instructions were presented and controlled by an IBM-compatible computer.

**Materials and Design.** Following the procedure of the previous experiments, 16 high and 16 low frequency lists, of eight items each, were created.

Sixty-four additional words of each frequency were randomly chosen to serve as distractor items for the recognition tests. Test ordering (recognition versus position) was blocked, and the experiment was counterbalanced such that each list appeared an equal number of times across all combinations of test type and test order (i.e. recognition first, position judgement second; position judgement first, recognition second). Within a block, presentation of the high and low frequency lists was also randomly determined for each subject.

**Procedure.** This experiment followed the procedures of the previous experiments in all respects other than testing. Within a block, subjects received eight high and eight low frequency lists followed by a single test type. For the recognition tests, subjects were shown a random ordering of the eight target and eight distractor items. These words were presented one at a time and remained in the centre of the terminal screen until a response was made. Subjects were asked to call those words that occurred in the studied list "old". They indicated whether each word was "old" or "new" by pressing, respectively, the number 1 or 2 key on the computer keyboard. Subjects initiated the start of each new list by pressing the space bar.

For the position judgement tasks, subjects were shown all eight list items, one at a time, in the centre of the terminal screen. Their task was to indicate which position each item held in its original presentation. Subjects were instructed to press the number key (1-8) corresponding to their position judgement. Subjects initiated the start of each new list by pressing the space bar.

Both the recognition and position judgement blocks were preceded by two practice trials; subjects were not told that the task would change later in the experiment.

## Results and Discussion

**Recognition.** The proportion correct responses (hits) at each serial position are shown in the top panel of Fig. 3. The  $2$  (frequency)  $\times 2$  (test order)  $\times 2$  (list)  $\times 8$  (serial position) ANOVA on these data revealed reliable effects of frequency,  $F(1, 40) = 25.25$  ( $MSE = 0.029$ ), serial position,  $F(7, 280) = 3.75$  ( $MSE = 0.012$ ), the frequency by serial position interaction,  $F(7, 280) = 3.55$  ( $MSE = 0.012$ ), and the three-way interaction of frequency, serial position, and list,  $F(7, 280) = 2.42$  ( $MSE = 0.012$ ).

Of most importance for the present concerns is the finding that the mean proportion of low frequency hits (0.91) significantly exceeded the proportion of high frequency hits (0.84). Of the 44 subjects in this experiment, 34 showed a low frequency advantage, eight showed a high frequency advantage, and there were two ties. This low frequency advantage is also evident in the analyses of

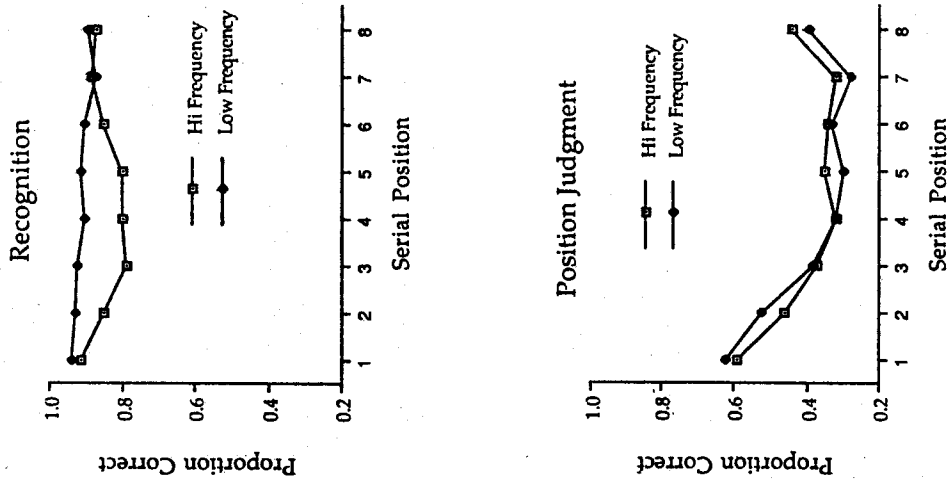


FIG. 3. Proportion correct recognition and position judgement responses as a function of word frequency and serial position.

$d'$ s<sup>1</sup>, which are shown in Table 1. A  $2$  (frequency)  $\times 2$  (order)  $\times 2$  (list) ANOVA on these data revealed only a reliable effect of frequency,  $F(1, 40) = 31.58$  ( $MSE = 0.15$ ).

The proportion of false alarm responses were analysed using a  $2$  (frequency)  $\times 2$  (order)  $\times 2$  (list) ANOVA. This analysis revealed reliable effects of frequency,  $F(1, 40) = 15.32$  ( $MSE = 0.015$ ), and test order,  $F(1, 40) = 6.69$  ( $MSE = 0.002$ ). The proportion of false alarms to high frequency words (0.06)

<sup>1</sup> Infinite values were avoided by applying the conversion suggested by Macmillan and Creelman (1991, p.10).