FREE RECALL: ORGANIZATION AND LONG-TERM MEMORY

by

ROBERT MARC SCHWARTZ
A.B. University of California, 1969
M.A., University of British Columbia, 1970

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

in the Department
of
Psychology

We accept this thesis as conforming to the
required standard

THE UNIVERSITY OF BRITISH COLUMBIA
July, 1972
In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the Head of my Department or by his representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Psychology

The University of British Columbia
Vancouver 8, Canada

Date August 3, 1972
ABSTRACT

Free recall was examined in order to (1) examine the differences between the clustering of items from a list with an a priori structure and the subjective organization from a list of unrelated words and to (2) describe organizational processes in terms of theoretical constructs.

Three sections of experiments found: (1) that two manipulations of organization of lists of unrelated words had asymmetric effects on recognition and recall performance; (2) that order of output and recall performance from two lists of unrelated words was predicted by similarity judgments data; and (3) that order of output from two specially-constructed lists was described by a hierarchical arrangement of categories.

The data were interpreted as showing three commonalities between the clustering of items from a list with an a priori structure and the subjective organization from a list of unrelated words. These were: (1) asymmetric effects on recognition and recall can be produced by manipulations of both categorized and unrelated lists; (2) order of output can be predicted by similarity judgments data for both categorized and unrelated lists; and (3) free recall performance can be predicted by conformity to category structures for both categorized and unrelated lists. An informal model of free recall was proposed. The model had the following properties which were abstracted from the experiments: (1) the internal structure of organization was described as hierarchical; (2) the processing which results in this structure was based on similarities between items; and (3) the output was described as representing different processes dependent upon whether recall or recognition tests are used.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A Review of the Data</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>A Review of the Theories</td>
<td>29</td>
</tr>
<tr>
<td>3</td>
<td>Organization: A Storage or Retrieval Effect?</td>
<td>48</td>
</tr>
<tr>
<td>4</td>
<td>Clustering and Subjective Organization</td>
<td>73</td>
</tr>
<tr>
<td>5</td>
<td>An Examination of Order of Output</td>
<td>100</td>
</tr>
<tr>
<td>6</td>
<td>Summary</td>
<td>127</td>
</tr>
<tr>
<td>7</td>
<td>Theoretical Notes</td>
<td>136</td>
</tr>
<tr>
<td>Table</td>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>-------</td>
<td>----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>I</td>
<td>Variances, Means and Transformed Means and Variances for the Four Conditions in Experiment I</td>
<td>56</td>
</tr>
<tr>
<td>II</td>
<td>ANOVA Source Table for Conditions in Experiment I</td>
<td>57</td>
</tr>
<tr>
<td>III</td>
<td>ANOVA Source Table for Imagery-Recall and Standard-Recall Conditions as a Function of Serial Position in Experiment I</td>
<td>60</td>
</tr>
<tr>
<td>IV</td>
<td>Graeco-Latin Square Design Used in Experiment II</td>
<td>64</td>
</tr>
<tr>
<td>V</td>
<td>Means and Variances and Transformed Means and Variances for the Four Conditions in Experiment II</td>
<td>66</td>
</tr>
<tr>
<td>VI</td>
<td>ANOVA Source Table for Conditions in Experiment II</td>
<td>67</td>
</tr>
<tr>
<td>VII</td>
<td>ANOVA Source Table for Silent-Recall and Overt-Recall Conditions as a Function of Serial Position in Experiment II</td>
<td>70</td>
</tr>
<tr>
<td>VIII</td>
<td>Selection Values as a Function of Number of Categories for the Cluster Analysis of the Sorting and Free Recall Data of Experiment IV</td>
<td>86</td>
</tr>
<tr>
<td>IX</td>
<td>Category Structures for the Sorting and the Free-Recall Data in Experiment IV</td>
<td>87</td>
</tr>
<tr>
<td>X</td>
<td>HGroup Selection Values for Group Data and Data for Two Individual Subjects as a Function of Number of Categories in Experiment VI</td>
<td>108</td>
</tr>
<tr>
<td>XI</td>
<td>Category Structure for Mammal Names Reported by Henley (1969) and Found in Free Recall for the Group Data and Data for Two Individual Subjects in Experiment VI</td>
<td>110</td>
</tr>
<tr>
<td>XII</td>
<td>Chi-squared Values and Contingency Coefficients for Relationships between Cluster Structures for Henley's (1959) Data and Group Data and Two Individual Subjects in Experiment VI</td>
<td>113</td>
</tr>
<tr>
<td>Figure</td>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>1</td>
<td>Probability of Recall as a Function of Serial Position for the Imagery-Recall and the Standard-Recall Conditions in Experiment I</td>
<td>59</td>
</tr>
<tr>
<td>2</td>
<td>Probability of Recall as a Function of Serial Position for the Silent-Recall and the Overt-Recall Conditions in Experiment II</td>
<td>69</td>
</tr>
<tr>
<td>3</td>
<td>Recall and Organization Scores as a Function of Trials for all Subjects in Experiment IV</td>
<td>89</td>
</tr>
<tr>
<td>4</td>
<td>Upper: Recall as a Function of Trials for High, Middle, and Low Organizers in Experiment IV Lower: Organization as a Function of Trials for High, Middle, and Low Organizers in Experiment IV</td>
<td>92</td>
</tr>
<tr>
<td>5</td>
<td>Recall and Z-Scores as a Function of Trials for Experiment VI</td>
<td>106</td>
</tr>
<tr>
<td>6</td>
<td>Hierarchical Category Structure for the Group Data in Experiment VI</td>
<td>111</td>
</tr>
<tr>
<td>7</td>
<td>Hierarchical Category Structure for Individual Subject 1 in Experiment VI</td>
<td>115</td>
</tr>
<tr>
<td>8</td>
<td>Hierarchical Category Structure for Individual Subject 2 in Experiment VI</td>
<td>116</td>
</tr>
<tr>
<td>9</td>
<td>Recall and Two-Way Subjective Organization Scores as a Function of Trials in Experiment VII</td>
<td>121</td>
</tr>
</tbody>
</table>
ACKNOWLEDGMENT

I thank the members of my dissertation committee for their careful criticisms of this manuscript. Dr. Michael S. Humphreys, my committee chairman, was instrumental throughout the preparation of this manuscript, and he deserves special thanks not only for help with this dissertation but also for providing a stimulating academic environment throughout my graduate career. I thank my friends, Duncan Hill and Keith Borkman, for help with preparation of the figures. I also thank my parents for moral and financial support throughout my education, and it is to them that I dedicate this dissertation.
CHAPTER 1

A REVIEW OF THE DATA

Methodologically, the free recall paradigm is perhaps the most simple of the verbal learning paradigms. In a typical free recall experiment the subject is presented a list of words, usually in a random order. He is then asked to recall as many of the words as he can, in any order that he desires. In the case of single-trial free recall, the experiment ends when the subject can recall no more words.

Unlike the single-trial free recall experiment, which may be classified as a memory paradigm, multitrial free recall can be regarded as a learning paradigm. In multitrial free recall, the subject is presented a list of words and is asked to recall them, just as in single-trial free recall. However, after the subject has finished recalling so many words as he can, he is again presented the list, usually in a different random order. After the second presentation, the subject is again asked to recall the words. This process of presentation followed by recall is continued for a fixed number of trials or until a preset criterion of performance is reached.

A third, albeit less frequently used, variation of the free recall paradigm involves multiple presentations of the words to be recalled but only one trial in which to recall them. That is, the subject is presented the list of words for a fixed number of trials; during these presentation trials he may be required to perform some operation involving the words (e.g., sorting). However, he is asked to recall the words only after the final presentation trial.
Although the free recall task is methodologically more simple than paired-associate or serial-learning tasks, free recall requires a different terminology. Traditionally, paired-associate and serial-learning tasks have been described in terms of stimuli and responses. For example, for a given response in paired-associate learning, the experimenter-defined or nominal stimulus is the stimulus term which has been paired with that response. In the case of rote paired-associate learning, the stimulus term is probably the effective stimulus, or the stimulus that the subject uses to produce the response. Even when the production of a response is said to be mediated by a learned association (cf., Kjeldergaard, 1968) or an image (cf. Paivio, 1969), the nominal stimulus must elicit the mediator which is the effective stimulus for the response.

In serial learning, the nominal stimulus is the word preceding the required response in the serial list. The effective stimulus may be identical to the nominal stimulus (e.g., Bugelski, 1950), may be a representation, elicited by the nominal stimulus, of the ordinal position of the required response (e.g., Mueller, 1970), or may be a combination of the two (e.g., Schwartz, 1970). In any case, it is clear that in serial learning, as in paired-associate learning, the nominal stimulus is easily identified and directly linked or identical to the effective stimulus.

In a free recall task in which the subject is allowed to recall the words in any order that he desires and no cue is presented to elicit a particular response, there is no nominal stimulus. If free recall is to be described in stimulus-response terminology, any concept of an effective stimulus must be a subject-produced stimulus. To describe a subject-
produced stimulus necessitates a theory about the nature of that stimulus and its formation. Thus, to describe free recall without subscribing to a particular theory requires a language other than that of stimuli and responses. One possible atheoretical terminology would describe free recall as consisting of input and output periods, referring to presentation and recall periods, respectively. Another possible terminology, which is somewhat theoretical, would describe free recall as a process of storing and retrieving the items in the list.

Closely aligned to the problem of the stimulus in free recall is the problem of order of output. As the subject is allowed to recall the words in any order, order of output is a subject-determined variable. This subject-determined variable would present little problem if subjects recalled the words in a random fashion or if their recall order corresponded directly with the presentation order. With randomly-ordered recall, it could be assumed that properties of individual items contribute to the "strengths" of the items and that the "strongest" items are recalled first. With presentation order recall, free recall might be described as analogous to serial learning. However, order of output is not random, nor does it directly correspond to the presentation order. Thus, unlike theories of paired-associate or serial learning, a theory of free recall demands a description of the order of output. This description may be independent of the description of the "stimulus" in free recall, or it may treat the problems of order of output and the stimulus as one problem.

Although lack of restriction on order of output increases the
complexity of describing free recall, it also increases the fascination of the paradigm. Obvious and interesting questions have yet to be answered satisfactorily. Why does the subject order his output? On what basis does he order his output? How can the structure of his output order be described? These questions provide the motivation for the current investigation; however, they cannot be answered in their present form. In order to operationalize these questions, it is worthwhile to review the free recall literature. Two reviews will be undertaken. In this chapter, the data and methodology concerning organization in free recall will be described. In the next chapter, theories which have been proposed to account for organization in free recall will be evaluated in terms of the data presented in this chapter.

Two Types of Organization

Two types of organization in free recall, clustering and subjective organization, have been identified. The major methodological differences are that clustering is experimenter-defined and may be measured in single-trial free recall, while subjective organization is subject-defined and can only be measured in multtrial free recall. These two types of organization will be discussed in separate sections.

Clustering

Bousfield and Sedgewick (1944) asked subjects to list items belonging to specific categories, e.g., birds, and found that many sequences of related items appeared in the subjects' listing protocols, e.g., duck and goose, hawk and eagle. Because he was unable to quantify
this "clustering" of related items in the listing protocols, Bousfield (1953) developed a method whereby the experimenter was able to control both the items to be listed and the relatedness of those items. He prepared a list of 60 two-syllable nouns; these were 15 nouns from each of four categories: animals, vegetables, names, and occupations. The nouns were placed in a random order and read to groups of subjects at a rate of one word every three seconds. After presentation was completed, the subjects were asked to list, in any order that they desired, as many of the words as they could remember. This procedure is, of course, identical to the procedure described as single-trial recall.

The measurement of clustering that Bousfield used involved the number of category repetitions in a given subject's recall protocol, i.e., the number of occurrences of a word from a given category followed by a word from the same category. In order to determine an "expected" number of repetitions to compare with each subject's observed number of repetitions, Bousfield ran an "artificial" experiment in which he drew capsules at random and without replacement from a box containing 15 blue, 15 green, 15 orange, and 15 white capsules. This procedure was repeated for each subject in the experiment, drawing a number of capsules that corresponded to the number of words that subject recalled (mean recall was about 25 words) and then counting the number of colour repetitions. Thus, Bousfield was able to compare the number of repetitions for subjects in his "real" experiment with those of his "artificial" experiment; he found that his "real" subjects clustered more than his "artificial" subjects.
A second analysis involved measuring changes in the clustering tendency during the course of recall. To do this, Bousfield "Vincentized" the subjects' recall protocols; that is, he divided each subject's recall protocol into successive decile intervals in terms of the order in which the items had been recalled. He found that the clustering tendency was initially above chance (chance was defined as the results of the artificial experiment), rose to a maximum at about the fourth decile, and progressively dropped to the chance level by the tenth decile.

Bousfield's (1953) experiment was concerned with one type of conformity of subject's order of output to an a priori structure of the list to be recalled, i.e., the clustering of members of a taxonomic category. Since Bousfield's original experiment, a variety of other types of clustering has been reported. For example, Jenkins and Russell (1952) found clustering of direct associates as measured by word association norms, and Tulving (1962a) found clustering of words having the same initial letter. There is no need to review all experiments identifying new types of clustering as both Bousfield and Bousfield (1966) and Shuell (1969) have made partial listings of such experiments.

One further point should be made. Most experiments designed to provide evidence for clustering in accordance with an a priori structure have used a single-trial free recall paradigm. However, in some cases, clustering has been examined as a function of trials in a multitrial free recall experiment (e.g., Puff, 1970). The typical finding is that clustering increases as a function of trials and that the correlation between clustering scores and recall scores over trials is very high.
Subjective Organization

There is substantial evidence that subjects' recall protocols conform with an a priori structure of the list to be recalled. Do subjects' recall protocols also show organization when the list to be recalled does not have an a priori structure?

Tulving (1962b) attempted to observe organization of a list of "unrelated" words, words chosen so that there are no systematic intra-list relationships apparent to the experimenter. He reasoned that in a list of unrelated words, there was no method for determining the degree of organization in single-trial free recall, so Tulving used multtrial free recall in order to observe the organization of a given subject. Organization scores were defined in terms of the number of trials on which the subject recalled word x followed by word y. That is, Tulving's measure of organization involved the number of trials on which two words appeared adjacently in the subject's recall protocols. This measure Tulving termed subjective organization, with subjective indicating that the more frequently occurring adjacencies may differ from subject to subject.

Tulving used a 16-word list and 16 presentation trials. The words were presented at a rate of one word per second, and after each presentation, the subjects were allowed 90 seconds to record their recall. As did Bousfield, Tulving used artificial subjects whose performance was matched with his real subjects; these artificial subjects provided a baseline for determining the degree of organization used by his real subjects.
Tulving's experiment yielded four main results. First, overall organization scores for his real subjects were far greater than for his artificial subjects. Second, organization scores and recall scores increased over trials. The correlation between the mean recall and log mean organization scores as a function of trials was .96. Third, total organization scores and total recall scores over subjects were found to be significantly correlated, $r = .63$. That is, subjects who showed high levels of organization tended to recall more than did subjects who showed low levels of organization. Fourth, there was a certain degree of commonality among subjects in the more frequently occurring adjacencies. This commonality was shown to increase as a function of trials.

Tulving's results have frequently been replicated, even in experiments in which different measures of organization have been used (e.g., Bousfield, Puff, & Cowan, 1964). It seems safe to conclude that correlation between organization measures and recall performance have a high degree of generality.

At this point, two types of organization have been discussed. For the remainder of this paper, organization of a list of words with an a priori structure will be referred to as clustering, and organization of a list of words with no a priori structure will be referred to as subjective organization.

Measurement of Organization

Until this section, organization has been defined on a fairly abstract level. However, operational definitions of organization do
exist in terms of the indices that have been proposed for both clustering and subjective organization. In this section, the indices that are used in the experiments to be reported will be discussed. For a review of other indices, the reader is referred to recent papers by Shihll (1969) and Frankel and Cole (1971).

**Measurement of Clustering**

For measurement of clustering, Frankel and Cole's (1971) adaptation of the multiple-category runs test was used. The important statistic for this test is the number of runs of items from the same category that appears in a subject's recall protocol. For example, a hypothetical subject recalls eight items in the following order: vegetable, vegetable, vegetable, animal, occupation, occupation, vegetable, name. The number of runs in this string is five, i.e., one more than the number of times that the category of the recalled item was changed. Wallis and Roberts (1957) have presented formulae for the expected mean and variance of the multiple-category runs test, and Frankel and Cole have described these formulae in terms of category clustering. As the mean number of runs is normally distributed, knowledge of the expected mean and variance and the observed mean can lead to the calculation of a $z$ score. When a $z$ statistic indicates that the number of observed runs was significantly fewer than the number of expected runs, evidence for clustering can be said to have been found. Frankel and Cole recommend that the $z$ score be used as a measure of the magnitude of clustering because a $z$ score is inversely related to the probability that the observed clustering is due to chance.
The runs measure presents the following advantages (Frankel & Cole, 1971):

1. the distribution of the z statistic is known, so that for a given subject, significance of clustering can be determined;
2. the distribution of a set of z scores is known, so that for a group of subjects, the significance of clustering can be determined (by $\sqrt{n} \times \bar{z}$ where n is the number of subjects);
3. the z statistic is independent of the number of items recalled, so that any observed relationship between z scores and recall performance will not result from an artifact of the clustering measure;
4. the categorical composition of the recall protocol is incorporated into the determination of the expected mean and variance, so that a given subject's z score is independent of the particular items that he recalls.

Frankel and Cole have reviewed other measures of clustering and found that the z statistic is the only measure which exhibits all four of the preceding advantages.

Measurement of Subjective Organization

As Shuell (1969) has pointed out, various indices of subjective organization are much the same; they all involve the number of intertrial repetitions and differ mainly in their corrections for chance organization. As all of the measures are highly correlated and, unlike the runs test, none of the measures has a distribution which is known, choice of a measure for the present experiments was somewhat arbitrary. The Bousfield and Bousfield (1966) deviation measure was chosen because it
is the most frequently used measure of subjective organization and
because a computer programme was available for its calculation.

For measurement of subjective organization, Bousfield and Bousfield
defined an observed intertrial repetition (OITR) as the case where word y
follows word x on both trial \( n \) and trial \( n+1 \). This measure can be ex­tended by defining an OITR as the case where word y follows word x on
trial \( n \) and either word y follows word x or word x follows word y on
trial \( n+1 \). It is this extension to a two-way index that will be used
in the present paper.

Bousfield and Bousfield also present formulae for determining the
expected number of intertrial repetitions (EITR). The formulae involve
the following parameters: (1) \( h \), the number of items recalled on trial
\( n \); (2) \( k \), the number of items recalled on trial \( n+1 \); and (3) \( c \) the
number of items which were recalled on trial \( n \) and also recalled on trial
\( n+1 \). Given an observed and an expected number of intertrial repetitions,
the measure of organization is simply, OITR-EITR. It is apparent that
this measure involves only the deviation of the observed from the ex­pected value and, unlike the runs test, does not take into account the
expected variance. Thus, no test for the significance of organization
of a single protocol exists, and the resulting values may not be com­parable for lists of different lengths.

Comparisons of the Two Types of Organization

As measures of clustering are not equivalent to measures of sub­jective organization, direct comparisons of the two types of organization
are difficult. The failure to have equivalent measures results from the
fact that given the recall of word $x$, adjacent recall of only one particular word (the word adjacent to word $x$ on the preceding trial) will lead to an increment of a subjective organization score, while adjacent recall of a number of words (words which are members of the same category as word $x$) will lead to an increment of a clustering score.

One attempt to compare the two types of organization was reported by Puff (1970a). Puff presented subjects a list of either 18 unrelated words or a list containing six items from each of three taxonomic categories; the words were presented for fifteen trials at a rate of 2.5 seconds per word, and one minute was allowed for each recall period. Although syllable length and Thorndike-Lorge (1944) frequency were similar for both lists, Puff found that the categorized list was easier to learn than the unrelated list.

Puff's comparisons of the two types of organization involved the expression of both types as proportions of the maximum possible organization score. His findings showed that clustering scores in the categorized list were far greater than subjective organization scores of that list and that subjective organization scores for the categorized list were not significantly different from subjective organization scores for the unrelated list. These findings are open to three possible interpretations. One interpretation is that clustering is a stronger phenomenon than subjective organization. A second interpretation is that there is as much subjective organization in a list of related words as there is in a list of unrelated words. A third interpretation is that subjective units contain more than two words and that there is
little organization within a subjective unit.

The fact that none or all of these interpretations may be correct points to the difficulty involved in directly comparing the two types of organization. Perhaps the only reliable method of comparing clustering (of categorized words) to subjective organization (of unrelated words) would be to compare the effects of manipulations of independent variables on the recall of categorized and unrelated lists. Four such comparisons follow.

Order of Input

A number of studies have shown that when a categorized list is presented in a blocked fashion, so that all members of one category are presented before all members of another category, recall is superior to random presentation (e.g., Weingartner, 1964; Cofer, Bruce, & Reicher, 1966). Of course, no direct analogue of blocked presentation can be found for the case of an unrelated list. However, Tulving (1965), capitalizing on the relatedness of his "unrelated" words, has shown that presentation order can affect recall performance of an uncategorized list. He arranged two presentation orders of a list used in a previous investigation (Tulving, 1962b). In one of the presentation orders, adjacent pairs had been frequently recalled adjacently in his earlier study. In the other presentation order, adjacent pairs had been infrequently recalled adjacently in his earlier study. Multitrial free recall was used, and subjects were run to a criterion of one perfect recitation. Median trials to criterion for the group with frequently occurring adjacencies was 3.4, and median trials to criterion for the group with
infrequently occurring adjacencies was 7.0; the difference was statistically significant.

Thus, presentation order seems to affect the amount recalled in lists of words with an \textit{a priori} structure and lists of words with no \textit{a priori} structure; in both types of lists, adjacencies in presentation order can be arranged so that recall performance is improved.

**Presentation Rate**

Cofer, Bruce, and Reicher (1966) have shown that when the stimuli are categorized words, amount recalled and clustering increase with slower presentation rates. Shapiro and Ponce (1970), in an experiment using multtrial free recall of a list of unrelated words, found greater subjective organization scores and increased recall performance with slower presentation rates. Thus, for both a list of categorized words and a list of unrelated words, organization measures and recall performance are related to presentation rate.

**Cueing**

After having been presented a list of categorized words, subjects may be given the category names as recall aids. This procedure, \textit{cueing}, has sometimes been found to improve recall performance. When measures are taken to insure that subjects use the category structure, by providing the category names alongside the words during presentation (Tulving & Pearlstone, 1966) or by using blocked category presentation (Wood, 1969), cueing has almost always been found to improve recall performance. Tulving and Pearlstone (1966) provided category names at presentation and,
for subjects in the "cueing" groups, also at recall. They used lists of 12, 24, and 48 words in length and each list contained 1, 2, or 4 words per category. In all but one condition, cueing was found to significantly increase recall and even in that condition (a 12-item list with 4 words per category), cueing resulted in slightly better recall performance.

Recall of categorized lists may be analyzed into two components, recall of a category and recall of words within a category. Tulving and Pearlstone reported data on both of these components, and they used recall of one or more words from a category as their definition of category recall. Subjects in cueing conditions recalled significantly more categories than did subjects in control conditions. However, once a category was recalled, the number of items recalled within that category was not different for cued and control subjects. Thus the apparent effect of cueing is to allow subjects to recall more categories.

To examine the effect of cueing of subject-determined categories, Dong & Kintsch (1968) used a variation of a paradigm first used by Mandler (Mandler, 1967; Mandler & Pearlstone, 1965). In their experiment, Dong and Kintsch had each subject sort three lists, each containing 25 randomly-chosen words. For each list, the subject sorted the words into between two and seven categories; number of categories and basis of category organization were subject-determined. Subjects sorted each list until they reached a criterion of two consecutive identical sorts. After having reached the sorting criterion on a given list, subjects in two groups were required to provide names for their categories.
After having provided names for their third list, subjects were asked to recall as many of the words from the three lists as they could. During recall, subjects in the cueing (Group RC) condition were provided with their category names, while subjects in the control (Group NC) condition were not provided with their category names. The results indicated that cued subjects recalled more than did control subjects. As Tulving and Pearlstone found for categorized lists, the advantage of cueing was in the number of categories recalled; cued subjects recalled more categories than did control subjects but did not recall more words per category. Thus, cueing seems to have the same effect on both experimenter-determined and subject-determined categories.

**Number of Categories**

Examinations of the effect of number of categories of a list with an *a priori* structure on free recall performance presents two problems. First, as was the case with cueing, it is important that the subject use the category structure provided by the experimenter. Second, in lists of a fixed length, the number of items per category is inversely related to the number of categories. This perfect confounding of items per category and number of categories presents many difficulties in interpretation.

Tulving and Pearlstone's (1966) finding that the advantage of cued recall is to increase the number of categories recalled but not to increase recall within a category suggests that when number of items per category is held constant, the more categories the subject recalls, the more words he will recall. Thus, number of categories would appear to be
an important determinant of recall performance. In Tulving and Pearlstone's experiment, subjects who were presented lists of 48, as opposed to 24, words did not differ in items per category recalled but did differ in recall performance. Obviously, the advantage of subjects receiving the longer list was in the number of categories recalled. Within a list of a given length, under conditions of cueing, Tulving and Pearlstone found that subjects who were presented more categories recalled more words. Again, number of categories appears to be an important determinant of recall performance.

However, both of Tulving and Pearlstone's findings involving number of categories are questionable. First, although it is interesting that recall within categories did not differ between 24- and 48-word lists, it is not in the least surprising that subjects who were presented the 48-word list recalled more. Furthermore, in the Tulving and Pearlstone experiment, subjects who were presented 12-word lists recalled more items per category than did subjects who were presented 24- or 48-word lists. Second, in the control conditions in which no cue was presented during recall, the relationship between number of categories and recall performance was not a direct function. In fact, Tulving and Pearlstone's data indicate that, in the control conditions, recall performance may have been an inverse function of the number of categories. The exact relationship between number of categories and recall performance appears to be a function of list length. Also, it seems as if the control conditions may be more representative of free recall than are the cueing conditions. For instance, in the 24-item list, cueing
resulted in highest recall when there was only one item per category. This, of course, would be an unrelated list, and recall of an unrelated list has often been shown to be worse than recall of a categorized list (e.g., Tulving and Pearlstone's 24-item lists containing 12 or 6 categories). Thus, although it appears that the number of categories has an effect on recall performance, the effect is confounded with list length, items per category, and cueing.

Mandler (1967) has presented an extension of the effect of number of categories to a list of unrelated words. In a typical experiment, subjects sort a list of either 52 or 100 words into categories; they are usually allowed to choose the number of categories they use with the restriction that they use no fewer than two and no more than seven categories. The sorting process is continued until stable categorization is reached. Then, the subjects are asked to recall as many of the words as they can. Generally, a linear relationship between number of categories used and amount recalled is obtained.

Mandler's results suggest that number of subjective categories is directly related to recall performance. However, there are two possible methodological artifacts in Mandler's paradigm. First, sorting is self-paced and, as it is continued to criterion, number of sorting trials is not experimentally controlled. Second, as subjects are allowed to choose the number of categories they use, number of categories is not experimentally controlled. Mandler (1968) has stated that he has controlled for these possible artifacts. First, he has shown that the category-recall relationship is obtained when sorting is experimenter-paced and continued for a fixed number of trials (Mandler, 1967, Experiment H). Second, he
has shown that when subjects are assigned a number of categories in a between-subjects (Mandler, 1967, Experiment B) or a within-subjects (Mandler, 1968) design, the relationship is still found. However, Mandler did not assign number of categories to his subjects in his experimenter-paced task, nor did he control time per trial and number of trials in the studies in which subjects were assigned number of categories.

Schwartz and Humphreys (1972a) have assigned subjects number of categories in an experimenter-paced task, and they failed to find a significant relationship between number of categories and recall performance. In addition, both Schwartz and Humphreys (1972a) and Nelson, McRae, and Sturges (1970) have found only small relationships between number of categories and recall performance in experimenter-paced tasks in which subjects were allowed to choose their number of categories. Thus, Mandler's results, much like Tulving and Pearlstone's, can be severely questioned on methodological grounds. It must be concluded that both the number of taxonomic and the number of subjective categories can influence recall performance, but in neither case, can the direction or amount of influence be clearly predicted.

Miscellaneous

In the previous sections of this review, two types of organization were defined, their measurement was discussed, and the two types were compared. For purposes of this paper, there remains three other areas for review. Unfortunately, these do not fit quite so nicely into the previous format. The three sections include (1) inhibition, (2) individual differences, and (3) short- and long-term memory components of recall.
Inhibition

Until this section, the review has only dealt with studies in which subjects have learned one list. However, as with other types of learning paradigms, free recall has been studied as a function of preceding (proactive inhibition) and intervening (retroactive inhibition) lists.

Proactive inhibition. In examining the effects of proactive inhibition, this review considers only the case in which some or all of the words from the first list are contained in the second list. Three types of proactive inhibition paradigms include an overlap of first- and second-list items. In part-whole free recall, all of the first-list items plus some additional items are contained in the second list. In whole-part free recall, all of the second-list items are contained in the first list, but not all of the first-list items are contained in the second list. In whole-whole free recall, the two lists are identical. For all three paradigms, the control group learns two lists which have no overlap of items.

Tulving (1966), in two part-whole experiments, reported three main findings: (1) there was no difference between experimental and control subjects in total second-list recall; (2) there was an interaction between groups and trials on the second list, such that control subjects learned the second list at a faster rate than experimental subjects; and (3) on later trials of second-list learning, control subjects were performing slightly, but not significantly, better than experimental subjects. Tulving's results have been replicated by Novinski (1969) and
Since Tulving's (1966) original experiment, a number of factors have been shown to lead to improved whole-list performance in the part-whole paradigm. First, Wood and Clark (1968) have shown that subjects' whole-list performance increases when they are instructed about the relationship between the part list and the whole list. Second, Ornstein (1970) has shown that blocking part-list and new words during presentation leads to better whole-list performance. Third, Birnbaum (1968) has shown that whole-list performance is improved when the part list is composed of categories which are taken intact from the whole list. Fourth, Schwartz and Humphreys (1972b) have shown that when subjective organization is measured on the part list, high organizers recall more whole-list items than do low organizers. Schwartz and Humphreys have pointed out that in all four of the preceding studies, improved whole-list recall performance was correlated with an increased tendency to cluster part-list words during whole-list learning.

The data on both whole-part and whole-whole free recall are similar to those on part-whole free recall. Tulving and Osler (1967), in a whole-part free recall experiment, found that experimental subjects perform worse, in terms of total recall, than do subjects in an appropriate control group. Although the experimental subjects have an initial advantage, they are shortly (in terms of trials) at a disadvantage which remains throughout second-list learning. Novinski (1969) has replicated these findings. Tulving (1966) gave experimental subjects incidental learning of a list of words while control subjects performed an unrelated
task. Both groups were then given intentional learning of the same list. Tulving found that the experimental subjects had no advantage in terms of total recall, during intentional learning. In fact, Tulving's graphs for experimental and control group performance suggest the presence of a cross-over effect. That is, incidental learners had a small advantage on the first trial, and control subjects had a small advantage on the later trials. Tulving did not report any statistics concerning his Groups X Trials interaction. However, Schwartz and Humphreys (1972b) gave subjects two trials of intentional learning on a list of unrelated words, then ten trials of intentional learning on the same list. Subjects in the experimental group had a significant first-trial advantage, but did not recall significantly more words over the ten trials than the control groups. Also, the Groups X Trials interaction was significant, and control subjects had a slight, but non-significant, advantage on the later trials.

In summary, the work on proactive inhibition with overlap between first- and second-list words appears to be quite similar. Learning of a list containing elements of a previously-learned list gives subjects an initial advantage. However, control groups which have no overlap of items in the two lists generally recall slightly more than the experimental groups on later trials. Usually, there is no difference between experimental and control groups in total recall.

Retroactive Inhibition. Retroactive inhibition in free recall has been less frequently studied than has proactive inhibition. However, three
recent studies (Shuell, 1968; Wood, 1970; Tulving & Psotka, 1971) have examined retroactive inhibition. Shuell (1968) had experimental subjects learn two categorized lists. For the "same" experimental condition, both lists contained the same conceptual categories; for the "different" experimental conditions, the two lists contained different conceptual categories. Control subjects learned only one list. First-list recall performance, after interpolated second-list learning, was the main independent variable. The results indicated that control subjects recalled more than did subjects in both experimental groups, and "different" experimental subjects recalled more than "same" experimental subjects. The superior performance of control as compared to experimental subjects was due to both a greater loss of category recall and loss of items within categories for the experimental subjects. The superior performance of "different" experimental as compared to "same" experimental subjects was due to a greater loss of category recall, but not items within categories, for the "same" experimental subjects.

Wood (1970), in three experiments, manipulated the similarity of organization between the two lists by instructing subjects how to organize the lists. He found greater retroactive effects when his instructions increased the similarity of organization between the two lists.

Tulving and Psotka (1971) studied retroactive inhibition by having subjects learn and subsequently recall from one to six categorized lists. In all instances, each list contained different categories than any of the preceding lists. They found: (1) that the number of words recalled was an inverse function of the number of intervening lists;
(2) that most of the loss as a function of intervening lists was due to loss of category recall -- contrary to Shuell's (1968) findings, there was no apparent effect of intervening lists on recall within categories; and (3) that cueing with category names eliminated retroactive effects.

The three studies on retroactive inhibition suggest the following generalizations: (1) that the retroactive effect may be on the number of categories recalled rather than on the number of words recalled from within a category; (2) that retroaction is a direct function of the similarity of organization between the lists; and (3) that retroaction increases as a function of the number of intervening lists.

**Individual Differences**

There is evidence that differences between subjects in organization scores are consistent across lists. Gorfein, Blair, and Rowland (1968) had subjects learn four lists, each of a different type of stimulus material -- consonant trigrams, nonsense syllables, unrelated words, and categorized words. They then correlated the subjects' organization scores between each of the four lists to produce six correlations. All of the correlations were positive and the magnitude of a given coefficient was a direct function of the similarity between the stimulus materials. In an unpublished study, the present author had nine subjects learn two lists of unrelated words; the type of stimulus material was felt to be more similar to the type used in studies of subjective organization than was the material used by Gorfein et al. The rank-order correlation between organization scores on each of the two lists was significant, $\rho = .62, p < .05$, one-tailed.
As differences in organization scores are consistent, these scores may represent individual differences in verbal learning ability. Earhard (Earhard, 1967; Earhard & Endicott, 1969) has shown that subjective organization scores can be used to predict performance on both serial and paired-associate lists. The general procedure used by Earhard is to have subjects learn a free recall list, measure organization scores on that list, and identify groups of high and low organizers. Then, the subjects learn a serial or paired-associate list. Earhard's general finding is that high organizers learn the second list faster than do low organizers. Gorfein (1971) has refined Earhard's findings by showing that when performance on the free recall list is controlled by partial correlation procedures, organization scores do not correlate with paired-associate performance. When organization scores are controlled in the same way, free recall performance significantly correlates with paired-associate performance. From these results it can be concluded that organization scores predict individual differences in other verbal learning tasks only to the extent that they reflect differences in free recall performance.

Schwartz and Humphreys (1972a) have shown that the number of categories a subject uses may also be a measure of individual differences in verbal learning ability. They showed that across sorting tasks, subjects were consistent in the number of categories they used. Then, they found some evidence for a relationship between the number of categories a subject chooses to use on one sorting task and the number of words he recalls from a completely different task.
In summary, the data concerning individual differences in subjective organization are far from conclusive. However, the data do warrant that extreme care be taken when drawing causal inferences from observed relationships between organization scores and recall performance.

**Short- and Long-Term Memory Components**

Single-trial free recall generally yields a bow-shaped serial position curve with greater recency than primacy effects. Recent studies have provided plausible explanations for the better recall of items presented at the ends of the list. The recency effect has been identified as a short-term memory component, in that it is affected by manipulations which have been shown to affect recall in the Brown-Peterson (Brown, 1954; Peterson & Peterson, 1959) paradigm. For instance, Kintsch and Buschke (1969) have shown that manipulations of acoustic, but not semantic, similarity affect the recency portion of the serial position curve. Glanzer and Cunitz (1966) have shown that 30 seconds of interpolated activity can eliminate the recency effect.

As Kintsch (1970) has indicated, the recency portion of recall may be independent of organization. He has used interpolated activity following list presentation in studies in which he desired to examine the structure of organization. Following Kintsch (1970), some of the experiments reported in the present paper also used interpolated activity following list presentation.

Explanation of the primacy effect has been more elusive than explanation of the recency effect; however, Rundus and Atkinson (1970) have presented evidence that the primacy portion may be due to the
extra opportunity for rehearsal of the earlier items. They had subjects rehearse aloud during presentation of the list to be recalled and found a striking parallel between number of rehearsals and probability of recall, except for the recency portion of the list.

Thus, there may be two components to single-trial free recall. The recency portion represents a short-term memory component which may be eliminated by sufficient interpolated activity. The primacy and middle portions probably represent a long-term memory component in which the probability of recall is related to rehearsal activity.

Summary

Two variations of the free recall paradigm, single-trial and multi-trial free recall, have been described. Two important features of free recall were identified. First, there is no nominal stimulus in free recall and second, organization may be observed in subject's recall protocols.

Two types of organization, clustering and subjective organization, have been described. Clustering was defined as conformity of recall output with an a priori structure, while subjective organization was defined as the trial-to-trial repetition of adjacencies in the output of unrelated words. Clustering and subjective organization have been compared as a function of variations of presentation order, presentation rate, cueing, and number of categories. First, presentation order can be arranged so that both structured ("clustered") and unrelated ("subjectively-organized") lists can be better recalled. Second, slower presentation rates increase recall performance and organization measures of both structured
and unrelated lists. Third, cueing with category names serves to increase the amount recalled from both structured and unrelated lists by increasing the number of categories, but not number of words within a category, recalled. Fourth, number of categories appears to be an important variable affecting recall performance of both structured and unrelated lists; however, because of methodological difficulties, in neither case can the nature of the effect be clearly determined.

Inhibition, individual differences, and short- and long-term memory components of recall have been discussed. Proactive inhibition was described in part-whole, whole-part, and whole-whole free recall. Learning of two lists which have an overlap of items does not seem to have a facilitative effect, but on later trials, it seems to have an inhibitory effect. Retroactive inhibition was found to be a direct function of similarity of organization of the interpolated lists and the number of interpolated lists. Also, most of the retroactive effect was found in category recall, not recall within categories. Experiments identifying organization scores as a measure of individual differences in verbal learning ability were described. Positive correlations were reported between a subject's organization score on one free recall list and his performance on another free recall list or on a paired-associate or serial list. Experiments identifying two components of free recall, a short-term memory component and a long-term memory component, were reviewed. A procedure for eliminating the short-term memory component was described.
CHAPTER TWO

A REVIEW OF THE THEORIES

In Chapter 1, the major experimental findings concerning organization and free recall were discussed. This chapter will consider the theories which have been proposed to account for these findings. The theories will be critically reviewed in terms of their ability to account for the data presented in Chapter 1.

Perhaps essential to understanding most of the theories is an oft-cited paper by George Miller (1956). In this paper, Miller was concerned with an integer, or actually a small range of integers, 7±2. Miller's concern was related to two types of experimental paradigms, psychophysical judgments and the digit span. The two paradigms will be discussed separately.

In one type of psychophysical task, the subject is presented a number of stimuli varying along one dimension. Each of the stimuli is assigned a label or name, and upon presentation of one of the stimuli, the subject must produce its label or name. Important to understanding the performance of a subject in a task of this type is the concept of "bits" of information. A bit represents the amount of information necessary to decide between two equally likely alternatives (a binary decision). Two bits, then, represents the information necessary to decide between four equally likely alternatives, while three bits represents the information necessary to decide between eight equally likely alternatives, etc. Thus, a subject's performance in a psychophysical task, such as that described above, may be measured in terms of bits of information.
Miller cited a number of experiments in which absolute judgments of unidimensional stimuli were made. He drew from experiments of auditory, gustatory, and visual discriminations and found the range of bits successfully processed (channel capacity) by subjects across these sensory modes was 1.6 to 3.9, representing from 3 to 15 items. What most impressed Miller about these findings was the small variance across experiments in the number of bits successfully processed; the mean number of bits was 2.6, which corresponds to a mean of 6.5 items, while the standard deviation was 0.6 bits, so that one standard deviation from the mean includes a range of 4 to 10 items. Hedging slightly, Miller concluded that performance in a unidimensional psychophysical task could be described in terms of discriminations between 7±2 stimuli.

How can a subject discriminate between more than seven stimuli? The obvious answer is to have the stimuli differ along more than one dimension. For instance, if there were two stimulus dimensions, it would be expected that subjects could discriminate between 49 stimuli, etc. Miller, also, has argued that increasing the number of dimensions should increase channel capacity. He cited a number of experiments which provide evidence for this point, but the experiments indicate that the function relating the number of items to the number of dimensions is not a simple multiplicative one; rather, the addition of stimulus dimensions adds to the channel capacity at a decreasing rate.

A second task which concerned Miller was the digit span task. In this task, the subject is presented a number of stimuli and is asked to repeat them in the order in which they were presented. For various types of stimuli, the span has been found to be about 7±2 items.
Miller pointed out that subjects remember about 7±2 stimuli in the digit span task regardless of the amount of information that the stimuli convey. For instance, one binary digit represents one bit of information, and one decimal digit represents over three bits of information. Yet, subjects remember 7±2 binary digits and 7±2 decimal digits. Thus, in the digit span information is not constant, while the capacity for the number of items that can be recalled appears to be constant.

One way to increase capacity in the digit span task is to increase the amount of information conveyed by each stimulus. For instance, Miller reports experiments in which subjects recode binary digits into octal digits and show a threefold increase in the amount of information recalled. This recoding process has been referred to as chunking.

Free recall theorists have drawn analogies between Miller's (1956) paper and free recall experiments in trying to determine how subjects can learn to recall more than seven items. Two analogies can be drawn, one to the psychophysical task and one to the digit span task.

The analogy to the psychophysical task requires that the words to be recalled can be described in terms of multiple dimensions. This assumption appears to be intuitively plausible because words have semantic, acoustic, syntactic, etc. properties. This analogy needs one of two further assumptions to describe output in free recall. First, the process may be passive such that words are coded along various dimensions at input and clustering is a product of some systematic search of the dimensions. Second, the process may be active such that words are chunked or coded along only a specified number of dimensions, and either
the search is limited or the recall of the code for a chunk precedes recall of the words within that chunk. This second assumption, then is a mixture of digit span chunking and psychophysical coding, and thus is not a direct analogy; however, it may specify a stimulus for the free recall of the items. The first assumption is perhaps a more direct analogy, but without additional assumptions does not solve the problem of the stimulus in free recall.

The second analogy has been drawn to the digit span task. Here, the subject is assumed to actively chunk the items. Recall of additional items may proceed by either adding information to a limited number of chunks, or by forming a hierarchical arrangement of chunks, subchunks, etc. This second analogy provides a stimulus in terms of a code for a chunk, but it presents three other problems. First, why does order of output correspond to the properties of the stimulus words, (e.g., Bousfield, 1953)? To answer this question, further assumptions concerning the relationship of chunking to some storage structure is necessary. Second, the analogy ignores possible distinctions between short- and long-term memory. Third, decoding in a free recall task is not analogous to decoding in a digit span task. For instance, if, in a digit span task, a subject decodes three binary digits into an octal digit, when he recall the octal digit, only one possible set of three binary digits will be suggested. That is, there is a one-to-one mapping between the new code and the original stimuli. In a free recall task, however, if the subject recodes three animal names into the code "animal," the recall of "animal" can suggest many more items than the three animal
names that were presented. Thus, if a theory of free recall is based on an analogy to Miller's discussion of the digit span, it must add assumptions about how the subject discriminates which items have been presented from those which have not been presented.

Having argued that analogies relating free recall to Miller's suggestions based on the "magical number seven" are rather limited, the remainder of this chapter will be devoted to three theories based on these analogies: (1) a theory based on a chunking process where there are a limited number of chunks, and improvement in recall is a function of increasing the amount of information per chunk; (2) a theory in which chunks are assumed to be hierarchically arranged; and (3) a theory in which multiple dimensions are used to describe free recall.

A Chunking Model -- Tulving (1968)

Tulving (1968) recognized both the limited capacity of immediate memory and the differences between paired-associate or serial learning and free recall learning. He raised two questions. First, how can a subject show increases in recall, as a function of practice, beyond the limitations of immediate memory? Second, how can interitem associations in free recall be described, and what is the relationship between these associations and free recall performance?

To answer these questions, Tulving (1968) described two types of units of analysis, E-units and S-units. E-units, or experimenter units, may be defined as the units that the experimenter uses in scoring the data. For instance, E-units may be the number of words recalled or, in the case of categorized lists, the number of categories or the number of
words within a category recalled. S-units, or subjective units, on the other hand, are the functional units employed by the memory system. Increases in recall, as measured in terms of E-units, are assumed to reflect the increasing size of S-units.

Tulving's distinction between E-units and S-units is a direct analogy to Miller's (1956) hypothesis concerning the limited capacity of immediate memory. If a subject can recall only a given number of items or E-units, he must chunk these items into larger items or S-units if he is to show increases in recall performance. Thus, increases in recall reflect the increasing size of S-units.

What is the nature of S-units? Tulving (1968) described two possible types of S-units. The first type of S-unit is based on item-to-item associations. That is, an item is organized into a given S-unit according to its strengths of association with other items in the S-unit. The second type of S-unit is composed of associations to a common mediator. That is, items that share the same superordinate category will be grouped into the same S-unit. Tulving points out that it is premature to distinguish between the two possible types of S-units. An item that is a direct associate to another item will most certainly share some superordinate category with that item. Conversely, items that are in the same superordinate category will have some associative strength.

Regardless of which type of associations describes S-units, Tulving contends that increases in associative strength within S-units leads to increases in recall. This contention, then, allows free recall to be described in a manner similar to paired-associate or serial learning
where it is possible to describe learning in terms of the strengthening of associations between stimuli and responses.

This brief sketch of Tulving's theory presents as systematic an account as can be drawn. Tulving prefers to modify or embellish his theory in order to describe the free recall data. A description of such modifications will be made in terms of the data discussed in the first chapter.

The Data

Tulving (1968) did not distinguish between mechanisms underlying clustering as opposed to subjective organization; rather, he considers the differences to lie solely in terms of experimental operations used to examine the two types of organization. Thus, his theoretical considerations of the free recall data apply to both clustering and subjective organization.

Variables which affect free recall performance are assumed to affect the subject's opportunity to discover relevant associations and to use these associations in forming S-units. Thus, presenting related words in contiguous input order should increase the subject's ability to perceive relevant associations, use these associations in forming S-units, and lead to increased recall. Similarly, slowing the presentation rate would give the subject more time to find relevant associations. In both cases, measures of organization as well as recall performance should show improvement.

To explain the effect of cueing and number of categories, Tulving (1968) distinguished between the "availability" and "accessibility" of
S-units. The terms "availability" and "accessibility," may be looked at as analogous to the terms "storage" and "retrieval," respectively. That is, an S-unit may be stored and be available to the subject, but because there is a limited capacity retrieval system, that S-unit may not be accessible. If the number of S-units exceeds the retrieval capacity, some S-units will fail to be recalled. Providing cues should allow the subject access to these S-units, and recall should show improvement. Accessibility, apparently, only applies to S-units, as the number of items recalled within an S-unit is independent of whether the subject recalls the S-unit or whether he must be cued to recall the S-unit.

Following the argument that recall of items within categories is independent of the number of categories recalled, number of categories should be an important determinant of recall performance. If number of categories exceeds retrieval capacity, recall performance should be inhibited. Subjects would have to combine or rearrange categories in order to reach perfect recall performance. Thus, a subject's recall should improve as a function of the number of categories until the point at which the retrieval capacity is reached.

Tulving's most impressive theoretical predictions involve the part-whole and whole-part paradigms, although his interpretation has been recently questioned (Schwartz & Humphreys, 1972b; Slamecka, Moore, & Carey, 1972). His predictions involve his assumptions about the relatedness of associations and the limited retrieval capacity. For part-whole free recall of unrelated words (Tulving, 1966), the subject is
assumed to form S-units for the part lists. These S-units, on the average, will be inappropriate for whole-list organization. Because the subject can only recall a limited number of S-units, he cannot simply maintain his part-list S-units and form new S-units for the new whole-list words. Thus, he will have to modify and reorganize his part-list S-units. As this reorganization is assumed to take time, the inhibitory effects of part-list learning should not occur until later in learning. A similar argument has been made for the whole-part paradigm (Tulving & Osler, 1967).

Tulving's (Tulving & Psotka, 1971) description of retroactive inhibition is not so clear as his description of proactive inhibition. Rather than explain why retroactive inhibition occurs, he uses his distinction between availability and accessibility of S-units to explain how retroactive inhibition occurs. That is, he states that retroaction should represent the inability to retrieve S-units rather than a loss of items from storage. If an S-unit from an early list is retrieved, there should be no loss of items within the S-unit. If cues are presented for S-units from an earlier list, there should be no decrement in recall. His data (Tulving & Psotka, 1971) support this point of view.

To the author's knowledge, Tulving has made no published statement concerning the relationship between organization scores and learning ability. However, as organization scores in free recall have been shown to be related to paired-associate and serial learning scores, a possible explanation is apparent. Serial learning and paired-associate learning can be described in terms of the strengthening of associations, and
Tulving has described free recall learning in terms of the strengthening of associations. Thus, a subject who had high organization scores would be assumed to be good at forming associations and should do well in paired-associate or serial learning.

Tulving does not agree that the bow-shaped serial position curve found in single-trial free recall represents two storage systems. He argues, instead, that the items presented at the ends of the list may be more accessible. The increased accessibility may be due to some auxiliary information about the end items, such as temporal dating or acoustic traces.

**Evaluation**

The flexibility of Tulving's theory is probably its greatest strength, and the flexibility makes the theory difficult to criticize in terms of the data. However, there are some possible inconsistencies between theory and data. For instance, Shapiro and Bell's (1970) finding that some subjects show increases in recall without corresponding increases in organization scores is difficult to reconcile with the proposed necessity of forming S-units. Schwartz and Humphreys' (1972b) finding that subjects who do not reorganize their first-list S-units perform better at part-whole free recall suggest that the hypothesized limit to the number of S-units which can be retrieved needs to be further investigated. Shuell's (1968) finding of a loss of items within categories as a function of retroactive inhibition is difficult to reconcile with the independence of category recall and recall of items within categories.
However, the main criticisms of Tulving's theory lie not in its inability to explain some aspects of the data but instead in the absence of certain mechanisms. For instance, items in S-units are said to be related. How is this "relatedness" extracted from long-term memory? There is a limited number of S-units which can be recalled. Is there a limit on the number of items which can be included in an S-unit? If so, how could the theory handle this limit? Other problems involve an absence of data. For instance, cueing is supposed to serve as a retrieval aid. Can it, then, be demonstrated that the effect of organization in noncued recall is a retrieval rather than a storage effect? It is assumed that the same processes underlie both clustering and subjective organization. Can this assumption be demonstrated? These are some of the questions that will be the concern of the experiments reported in the next three chapters.

A Hierarchical Model -- Mandler (1967)

Mandler (1967) has proposed a model of organization which is very similar to that offered by Tulving (1968). The main difference between the two models is the hierarchical arrangement of higher-order memory units in Mandler's model as opposed to the chunking in Tulving's model. Like Tulving, Mandler, by extension from Miller's (1956) argument concerning the limited capacity of immediate memory, hypothesized that organization is necessary for improving free recall performance. However, Mandler also provided an answer for a question raised in regard to Tulving's model. That is, is there a limit to the number of items contained in an S-unit? Mandler suggested that there is such a limit,
and if the number of items within an S-unit surpasses this limit, sub-
jects will form subordinate S-units. Thus, a hierarchical structure
will develop.

Mandler (1967) also proposed a numerical value for the limit to
the number of units, items within units, etc. — 5±2. The difference
between the 5±2 that Mandler proposed and the 7±2 offered by Miller
(1956) results from Mandler's recognition of the differences between
short- and long-term memory. Using the results of an experiment con-
ducted by Waugh and Norman (1965), Mandler estimated the limit of short-
term memory as 3±1. The remainder, 4±1, was offered as Mandler's es-
timate of the long-term memory component. His experiments on the cate-
gory-recall function (see Chapter 1) have led him to revise this esti-
mate to 5±2.

As Mandler's (1967) theory has not been extended to cover as many
aspects of the free recall data as has Tulving's (1968) theory, Mandler's
theory will not be evaluated in terms of the data. In any case, because
of the similarity between Mandler's approach and Tulving's approach, it
is likely that Mandler would handle many aspects of the data in the same
way that Tulving has. However, it should be pointed out that Schwartz
and Humphreys (1972a) have severely questioned, on methodological grounds,
Mandler's category-recall function (see Chapter 1).

Though Mandler has not presented a systematic account of the data,
he should be credited with two theoretical improvements of Tulving's
theory. First, Mandler presents an approach for handling a possible
limited capacity of items within S-units. Second, he distinguishes
between short- and long-term memory components in dealing with his analogy from Miller's (1956) review.

A Multidimensional Model — Kintsch (1970)

Kintsch (1970) has presented a model of free recall which may be classified as a multidimensional model. Kintsch's model, however, is not based on a direct analogy to the psychophysical judgment tasks discussed by Miller (1956), nor does the model directly refer to a multidimensional structure. Nevertheless, Kintsch does describe words in long-term memory as represented by a series of "markers." If these markers are allowed to assume continuous values, the series of markers for a given word would indicate its location in a multidimensional space.

Kintsch refers to three primary types of markers, S, I, and P. S markers are syntactic-semantic markers and are of greatest importance in free recall. I markers are sensory feature markers, and P markers are phonetic markers. S markers may be broken down into lexical, associative, and syntactic fields. Here, lexical fields are of the greatest importance. Lexical fields may be of two types. The first type, which Kintsch calls antonymy, is given by oppositions, e.g., good-bad. In some cases, these oppositions can be seen as extremes of a dimension. The second type of lexical field is called hierarchical or sequential. Markers in this type of field refer to relationships of class inclusion or category membership.

When a word is presented in a free recall task, the subject stores the information that the word has occurred. He does not store the word per se because the word is already present in long-term memory. Instead,
he tags the marker list that corresponds to the word in order to indicate that the word was presented.

To follow the course of input and output in free recall, Kintsch uses the Atkinson and Shiffrin (1968) "buffer" model. A word, when presented, has some probability of entering a short-term memory buffer (for variables which determine whether a word will enter the buffer, see Atkinson & Shiffrin, 1968). If the word does enter the buffer, some "cognitive" work will be performed on it. This cognitive work can be of two types. First, the word may be added to the marker list of another word in short-term memory. Second, and both more efficient and more in line with a multidimensional model, a marker may be found that is held in common with another word. The efficiency of this cognitive work will vary directly with the amount of time the item remains in the buffer.

In retrieval, items from the short-term memory buffer are first recalled. Then, another word is chosen in a random manner. The markers associated with this word are examined. If a marker for this word has been associated with another item, that item is examined. If that item's markers indicate that the item was on the list, the item is recalled. The process can then recycle with this new item. If no item is found which was on the list, the search stops or a new item is chosen to recycle the search.

A model of this type predicts that similarity of items should be the best predictor of output order in free recall. That is, the more similar two items are, the more likely they are to share common markers, and the "cognitive" work performed in the buffer should be more efficient
in finding these common markers. Thus, clustering of related items is readily predicted. Also, the greater recall of categorized lists, as opposed to unrelated lists, can be predicted as a function of the probability of the search stopping. That is, with unrelated lists, it is less likely that the set of markers associated with a recalled item will include a marker which is shared in common with another item on the list.

The Data

Kintsch's (1970) model is in its formative stages, and no systematic attempt has been made to use the model in explaining a large variety of free recall phenomena. In this section, an attempt will be made to show how a model of this type could handle some of the data discussed in the first chapter.

First, Kintsch does seem to differentiate between clustering and subjective organization. He refers to subjective organization as idiosyncratic (1970, p. 350). However, in his model, he does not indicate that the two types of organization represent different processes. Thus, in the analysis of the data, the two types of organization will be treated as if they were the same.

Order of input, presentation rate, cueing, and the serial position curve should be handled easily by Kintsch's model. Any manipulation which would allow the subject increased opportunity to discover common markers should result in increased organization and increased recall. Thus, when items which share common markers are presented adjacently, they are likely to reside in the buffer at the same time and the common
markers can be extracted. When the presentation rate is slowed, there is more time for the "cognitive" work which results in finding common markers. The search process could stop without certain markers having been found in the lists of markers that were examined. A cue word, then, could serve the purpose of a marker and aid retrieval of items associated with that marker. Kintsch's use of the Atkinson and Shiffrin (1968) buffer concept provides a mechanism for describing the serial position curve.

As Kintch's model does not include a concept of categories, it would be difficult to explain the effect of number of categories on free recall performance. Also, added assumptions would be necessary to explain inhibition and individual differences.

Evaluation

It is difficult to evaluate Kintsch's (1970) model in terms of the data because the model has not been applied to many aspects of the data. However, it is possible to compare Kintsch's model with a model such as that offered by Tulving (1968).

Kintsch's approach offers answers to questions of relatedness in recall protocols. Unlike Tulving, Kintsch suggested a structure for long-term memory and provided a mechanism for using that structure in free recall learning. However, the structure presents other problems. For instance, what prevents the subject from using most of his search time finding (but not recalling) items that were not presented? That is, a marker which describes an item which was presented should also describe many items which were not presented.
Kintsch also described subjective organization as being idiosyncratic while clustering was described as being based on similarities. The distinction between clustering and subjective organization was not made by Tulving. As yet, no data are available which examine subjective organization in terms of similarities.

Like Tulving, Kintsch described the effect of organization to be in retrieval rather than in storage. If an item enters the buffer, its marker list is tagged. Thus, whether or not an item is stored is a function of its having entered the buffer and not of organization. However, organization, in terms of common markers, is an important determinant of recall probability. Unlike the models offered by Tulving and Mandler, Kintsch's model does not imply that organization is necessary for improving free recall performance.

Conclusion

Because the three theories presented in this chapter differ in their degree of completion and the types of data they were formulated to explain, it would be difficult to choose the "better" of the three. However, there are differences between the theories, and some of these differences motivate the research presented in the next three chapters.

The major difference between the theories is that two of them (Tulving, 1968; Mandler, 1967) consider increases in organization necessary for increases in recall performance, and the third (Kintsch, 1970) considers increases in organization helpful for increases in recall performance. It was argued in Chapter 1 that it is impossible to distinguish between these alternative views of organization. However, it is possible to ask
a question related to these alternative views. That is, given that organization affects recall, at what stage of the recall process is the effect felt? Both Tulving (1968) and Kintsch (1970) argue that the effect is in the retrieval, rather than the storage, stage. However, analogies from Miller's (1956) notion of limited "channel" capacity do not necessitate a retrieval, rather than storage, effect, and there is little evidence which is directly related to the storage-retrieval question. In Chapter 3 two experiments directly related to the storage-retrieval question will be reported.

Another difference between the theories is that Tulving (1968) does not distinguish between clustering and subjective organization, but Kintsch (1970) may make such a distinction. Also, Kintsch's model has a mechanism for organizing according to similarity between items, but Tulving's model does not have such a mechanism. The experiments reported in Chapter 4 were designed to examine both of these differences. Here, Kintsch's notion of similarity was used to examine the organization of unrelated words. If order of output of unrelated words could be predicted from similarity judgments data, it could be argued (1) that the same processes underlie both clustering and subjective organization and (2) that a mechanism involving similarity can be used as an explanatory device for both types of organization.

The three theories also differ in the internal structures used to generate order of output. Tulving's (1968) model uses a chunking structure, and Mandler's (1967) model uses a hierarchical arrangement of chunks. Kintsch's (1970) model is more analogous to a 'cue-directed'
model of multidimensional search. In Chapter 5 order of output was examined in the hope that with an appropriate selection of materials, an adequate description of order of output might be made. Such description might reflect on the internal structure used to generate such output.

In Chapter 6, the data from Chapter 3, 4, and 5 were summarized and discussed. In Chapter 7, the data presented in the first chapter, the theoretical considerations raised in this chapter, and the data from the experiments reported in the next three chapters were incorporated into an informal model of free recall. Suggestions as to how the model could handle the data presented in Chapter 1 were made.
CHAPTER 3

ORGANIZATION: A STORAGE OR RETRIEVAL EFFECT?

Theories of free recall have generally involved the assumption that the effect of organizational processes is on the amount retrieved rather than the amount stored. For instance, Tulving (1968) differentiates between the "availability" and "accessibility" of items and argues that organization into S-units affects accessibility. His argument follows from his notion of a limited capacity retrieval system and from data he has collected on the effects of cueing (Tulving & Pearlstone, 1966) and retroactive inhibition (Tulving & Psotka, 1971). Kintsch (1970) also argues that organization affects retrieval rather than storage. Kintsch's argument is an extension of his two-process theory of recognition and recall. His model of recognition is based on the assumption that storage is determined by whether an item enters the short-term memory buffer where the probability of an item entering the buffer is not related to organizational processes. However, the probability of recall is dependent upon the extent of the "cognitive work" or organization that takes place in the buffer.

Although Tulving (1968) and Kintsch (1970) agree that organization affects the amount retrieved but not the amount stored, their theories do not eliminate the possibility of storage effects. For instance, the types of organization described by both Tulving (1968) and Kintsch (1970) affect the form of the stored information. Tulving views organization as analogous to the chunking process described by Miller (1956). Thus, several items in the list are recoded, and the code is stored. In
Kintsch's model, an item is represented by a series of markers which describe that item. The "cognitive work" performed on that item is represented by a change in its list of markers. For instance, an item may be added to the marker list of another item. Thus, in both Kintsch's and Tulving's models, organization affects the form of item storage.

As these theories do not preclude storage effects, the storage-retrieval question is a useful point of departure for an empirical investigation of organizational processes in free recall. In this chapter, two experiments which examine the storage-retrieval question are reported. The paradigm for both experiments is similar and takes the form of a 2 X 2 factorial design. The first factor is a manipulation of organization so that there are two groups of subjects who differ in the extent to which they have organized the material. The second factor is the form of testing, i.e., whether recognition or recall tests are used. It is assumed that recognition measures the amount stored and recall measures the amount retrieved. This assumption appears intuitively valid, and support for this assumption can be found in a paper by Kintsch (1970) and in a review article by McCormack (1972). If the organizational manipulation affects recognition performance, it can be concluded that organization does affect the amount stored. If the organizational manipulation affects recall performance but not recognition performance, it can be concluded that organization affects the amount retrieved but not the amount stored.

Both Kintsch (1968) and Earhard (personal communication) have examined recognition and recall performance as a function of differences in
organizational level. Kintsch's manipulation involved the type of material presented. All subjects were presented lists of 10 words belonging to each of four conceptual categories. For some subjects, the items were high associates to the category names; for other subjects, the items were low associates to the category names. Subjects were tested using either recognition or recall tests, and the results showed that subjects presented lists of high category associates recalled but did not recognize more than subjects presented lists of low category associates. Kintsch's results have been replicated by Bruce and Fagan (1971).

However, Kintsch's results cannot be clearly interpreted as indicating that organization affects recall performance but not recognition performance. First, to reach such a conclusion, it must be assumed that the only difference between presenting a list of high category associates as opposed to a list of low category associates is the amount of organization the subject uses. This assumption is not necessarily valid. Puff (1970b) has found that in single-trial free recall, subjects who clustered a categorized list more than would be expected by chance did not recall more than subjects who clustered the same list at the chance level, and that both groups of subjects recalled more than subjects who learned a list of unrelated words. Thus, it is not clear that the difference between recall of a list of high category associates and a list of low category associates is due to level of organization. Second, as the material presented varied between groups, recognition tests could not be equated. That is, the type of distractor used to test the recognition
of a list of low category associates is not equivalent to that used to
test the recognition of a list of high category associates.

Earhard (personal communication) has identified groups of high and
low subjective organizers and compared these subjects on recognition tests.
She found that high organizers performed better on the recognition tests
than low organizers. One possible conclusion from her results is that
organization affects the amount stored. However, as discussed in the
first chapter, organizational level may be a measure of some verbal
learning ability, and it is clear that Earhard's study does not separate
the degree of organization from this ability. Thus, a second possible
conclusion is that the locus of the individual ability which correlated
with organization scores is in the amount stored.

Experiment I

In the experiments reported in this chapter, an attempt was made
to avoid the confounding of organization with either type of material or
the subject's ability. In Experiment I this attempt was realized by
instructing subjects in the experimental groups to use visual imagery
to organize the items in the order in which they were presented. The
effect of instructions to use visual imagery in paradigms other than free
recall has been reviewed by Paivio (1969).

Insofar as "organization" may include the grouping of items at
input, instructions to use visual imagery in order to group items can
be interpreted as a manipulation of organization. Bower, Lesgold, and
Tieman (1969) have used imagery instructions to encourage subjects to group items; however, unlike a typical free recall situation, the Bower et al. experiments presented items in groups. In the present experiment, to create a more typical free recall situation, items were presented one at a time and experimental subjects were asked to use visual imagery to group items according to input order.

Subjects in the control groups were read standard instructions which did not mention visual imagery. Subjects in both experimental and control groups were given recognition and recall tests.

Method

Stimuli and materials. The stimuli were 120 concrete (C > 5,50) and frequent (F > 20) nouns chosen randomly and without replacement from the Paivio, Yuille, and Madigan (1969) norms. Of these 120 nouns, 60 were randomly chosen to comprise the list to be presented to all of the subjects. The other 60 nouns served as distractors for the subjects in the recognition conditions.

Each of the 60 nouns on the list to be presented was mounted as a slide. The slides were then randomly arranged to determine order of presentation for all of the subjects. Presentation was via a Kodak carousel projector (Model 850) with an external timer. Answer sheets were prepared for subjects in the recognition and recall conditions. For the recall conditions, the answer sheets contained two columns of 30 blank lines. At the top of the answer sheet was an instruction asking the subject to write down in any order he desired, as many words as he could remember from the list he had been presented. For the
recognition conditions, the 60 presentation items and the 60 distractor items were randomly arranged and typed in four columns of 30 items each. The instruction at the top of the answer sheet asked the subject to put a line through all of the words he thought were contained in the list he had been presented.

**Design.** A 2 x 2 factorial design, with both factors varying between groups of subjects, was used. The first factor was an instructional manipulation, i.e. whether the subject was given standard learning instructions or was given instructions to use visual imagery to organize the list. The second factor was type of test, i.e. whether the subject was given a recognition or recall test. Thus, there were four conditions which were labelled Standard-Recall, Standard-Recognition, Imagery-Recall, and Imagery-Recognition.

**Subjects.** The subjects were 57 volunteers from introductory psychology classes at the University of British Columbia. They were tested in groups of 1-3 with all subjects in a group receiving the same instructional manipulation. There were 14 subjects in each condition except the Standard-Recognition condition which had 15 subjects.

**Procedure.** Subjects in all conditions were told that they would be presented with 60 nouns that would be familiar to them. Upon seeing each noun, they were to try to remember it as best they could. They were also told that after seeing all 60 nouns, they would be tested on how well they had remembered them, although they were not informed of the nature of the test.

Subjects in the imagery condition were given additional instructions
about learning the list. Pilot work had indicated that these instructions were effective for increasing performance in a free recall task. The additional instructions follow:

Some students report that they can remember words of this sort better if they form visual images containing a number, usually five or fewer, of the words. This idea of forming visual images may best be explained by example. Suppose the first five words presented were monkey, tennis, shelf, sandwich, and rock. You might imagine two monkeys playing tennis and eating sandwiches on one shelf of a giant bookcase. Instead of a tennis ball, they were playing with a large rock. Then the next five words would be presented, and you would form another image with those five words and so on. Usually, the more unusual or bizarre an image you form, the easier it will be to remember the words.

We want you to try to use this method of forming visual images to help you remember the words. Try not to put more than five words in an image, because if you use more than five, you may find yourself confused. Remember, the words will always be presented at a constant rate; there will be no break after each set of five words to tell you when to stop forming one image and to start working on another. You will have to keep track of the number of words in each of your images on your own.

If the subjects had no questions, the list was then presented.

Rate of presentation was four seconds per item. Immediately following presentation and according to their conditions assignment, the subjects were handed either a recognition or a recall answer sheet. The subjects were told that the instructions were at the top of the sheet and they would have four minutes to complete the task.

Results

Following Kintsch (1968), the score for each subject was the difference between the number of correct and the number of incorrect responses. The means of these scores were 45.29, 42.27, 20.71, and 16.29 for the
Imagery-Recognition, Standard-Recognition, Imagery-Recall, and Standard-Recall conditions, respectively. Means and variances of these scores for each of the four conditions appear in Table I. To approximate homogeneity of variance, each subject's score was transformed using a square root transformation. The transformed means and variances are also shown in Table I.

The transformed data were subjected to analysis of variance. Table II contains a source table for that analysis. As Table II indicates, the imagery instruction resulted in significantly improved recall performance but not significantly improved recognition performance.

A second analysis was performed on the data of the recall conditions to assure that the superiority of the Imagery-Recall condition could be attributed to the instructional manipulation. As the subjects in the imagery conditions were told to form images of groups of five words according to the input order, it was predicted that the protocols of the subjects in the Imagery-Recall condition would show more clustering according to input order than those of the subjects in the Standard-Recall condition. To test this prediction, the presentation list was treated as if it contained 12 categories of five words each. Each category was a group of five words presented adjacently, starting with words 1-5, 6-10, etc. Then each subject's recall protocol was analyzed according to Frankel and Cole's (1971; see Chapter 1) adaptation of the runs test, and a $z$-score was determined for each subject. The mean $z$-scores for the Imagery-Recall and Standard-Recall conditions were 3.53 and 1.93, respectively. Both means indicated significantly more clustering according to
<table>
<thead>
<tr>
<th>Measure</th>
<th>Recognition</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{x}$</td>
<td>45.29</td>
<td>20.71</td>
</tr>
<tr>
<td>$s^2x$</td>
<td>98.38</td>
<td>47.91</td>
</tr>
<tr>
<td>$\sqrt{x}$</td>
<td>6.68</td>
<td>4.50</td>
</tr>
<tr>
<td>$s^2\sqrt{x}$</td>
<td>0.67</td>
<td>0.67</td>
</tr>
</tbody>
</table>
TABLE II
ANOVA Source Table for Conditions in Experiment I

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>ms</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>56</td>
<td>99.68</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Recognition vs. Recall</td>
<td>1</td>
<td>76.64</td>
<td>76.64</td>
<td>191.60</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Imagery Recognition vs. Standard Recognition</td>
<td>1</td>
<td>0.35</td>
<td>0.35</td>
<td>&lt;1</td>
<td>n.s.</td>
</tr>
<tr>
<td>Imagery-Recall vs. Standard Recall</td>
<td>1</td>
<td>1.72</td>
<td>1.72</td>
<td>4.30</td>
<td>&lt;.05</td>
</tr>
<tr>
<td>Error</td>
<td>53</td>
<td>20.97</td>
<td>0.40</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
input order than would be expected by chance (z = 13.20 for Imagery-Recall and z = 7.21 for Standard-Recall, both p's < .001). However, the Imagery-Recall condition showed significantly more clustering according to input order than the Standard-Recall condition, z = 4.24, p < .001. Thus, it can be concluded that the Imagery-Recall subjects were organizing in accordance with the instructions.

A third analysis examined the effect of the imagery instruction on recall as a function of serial position at input. Each subject's recall protocol was examined for probability of recall of items presented in serial position 1-15, 16-30, 31-45, and 46-60. The mean probabilities of recall as a function of these four blocks of input positions for the Imagery-Recall and Standard-Recall conditions are presented in Figure 1.

The data shown in Figure 1 were subjected to analysis of variance. The dependent measure was the number recalled at each of the four blocks of input positions for each subject. A source table for this analysis is contained in Table III. The results of this analysis indicate a significant effect of serial position but no significant Imagery-Recall vs. Standard-Recall X Serial Position interaction. Thus, the effect of the imagery instruction was not restricted to a certain portion of the serial position curve.

Experiment II

Experiment I indicated that the effect of the organizational manipulation was to increase the amount recalled but not the amount recognized. However, concluding that organization affects the retrieval
FIGURE 1

Probability of Recall as a Function of Serial Position for the Imagery-Recall and the Standard-Recall Conditions in Experiment I
TABLE III

ANOVA Source Table for Imagery-Recall and Standard-Recall Conditions as a Function of Serial Position in Experiment I

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>111</td>
<td>872.28</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Between Subjects</td>
<td>27</td>
<td>210.43</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Imagery-Recall vs.</td>
<td>1</td>
<td>31.08</td>
<td>31.08</td>
<td>4.51</td>
<td>&lt;.05</td>
</tr>
<tr>
<td>Standard-Recall (1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>26</td>
<td>179.35</td>
<td>6.89</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Within Subjects</td>
<td></td>
<td>692.93</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Serial Position (2)</td>
<td>3</td>
<td>202.39</td>
<td>67.46</td>
<td>10.93</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>1 X 2</td>
<td>3</td>
<td>9.02</td>
<td>3.01</td>
<td>&lt;1</td>
<td>n.s.</td>
</tr>
<tr>
<td>Error</td>
<td>78</td>
<td>481.92</td>
<td>6.17</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
capacity but not the storage capacity requires accepting the null hypothesis concerning the recognition data. In order to gain more confidence in accepting the null hypothesis, Experiment II was designed to replicate the results of Experiment I. In Experiment II, a different manipulation was used to create differences in organization. This manipulation, unlike the imagery instruction which resulted in higher organization, was intended to depress organization.

In his doctoral dissertation, Rundus (1970) had subjects rehearse aloud during list presentation and found that items which had been rehearsed together were likely to be recalled together. From this result, it can be inferred that subjects' rehearsal strategies are part of the organizational process. Thus, if the subject is instructed to rehearse aloud only the item being presented (Fischler, Rundus, & Atkinson, 1970), it is possible that his level of organization would be depressed. In Experiment II, subjects in the experimental group were required to rehearse aloud the item which was being presented. Subjects in the control group were not required to rehearse aloud and were not asked to use any particular rehearsal strategy. If the effect of organization is on retrieval capacity, it could be predicted that subjects required to rehearse aloud would recall less than control subjects but would not differ from control subjects in recognition performance.

**Method**

**Stimuli and Materials.** The stimuli were 480 words, three from each of 160 conceptual categories in the Battig and Montague (1968) and Stapiro and Palermo (1969) norms. The words were high associates to the category
names. One word from each of the categories was chosen to be on the lists to be presented; the remaining two words from each category served as distractors for the recognition conditions.

The 160 words to be presented were typed on 3 X 5 inch notecards. The notecards were then randomly divided into four lists of 40 words each. The lists were arbitrarily designated Lists A, B, C, and D. Within each list, the notecards were randomly arranged to form a presentation order for that list.

Recall answer sheets were prepared in the same manner as in Experiment I, except there were only 20 blank lines in each of two columns. Four separate recognition answer sheets were prepared, one for each of the four lists. A recognition answer sheet for a given list contained a random arrangement of the 40 words on that list and the two distractor words from the categories of those 40 words. The recognition answer sheets were arranged in four columns of 30 words each, and the instructions at the top of the recognition answer sheets were the same as those used in Experiment I.

**Design.** As Experiment I, Experiment II could be conceptualized as a 2 X 2 factorial design. The first factor would be rehearsal conditions, silent or overt, and the second factor would be type of test, recognition or recall. Thus, there would be four conditions, Silent-Recognition, Overt-Recognition, Silent-Recall, and Overt-Recall.

However, unlike Experiment I, the four conditions were varied within subjects so that each subject served in each of the four conditions. Each subject was presented each of the four lists and served in
one of the four conditions on each list. To counterbalance order of conditions and presentation of lists, a Graeco-Latin square design was used. The exact Graeco-Latin square is illustrated in Table IV. As can be seen from Table IV, the Graeco-Latin square resulted in four arrangements of lists and conditions for presentation.

Subjects. The subjects were 20 volunteers from introductory psychology classes. Five subjects served in each of the four Graeco-Latin square arrangements. The subjects were tested individually.

Procedure. Each subject was read two sets of instructions, one set before he served in each of the two rehearsal conditions. In neither set of instructions was the subject informed of the type of test which would follow the presentation of the list.

For both sets of instructions, the subject was told that he would be presented 40 words that were familiar to him. He was shown a deck of notecards on which one of the lists of words was typed and was told that he would be presented the list, one word at a time, by the experimenter turning over the notecards. The subject was told that a card would be turned over every five seconds and that he should try to remember as many words as he could. In addition, for the silent rehearsal condition, the subject was told he could study the words in any way that he wished. For the overt rehearsal condition, the subject was told to repeat aloud the word he was currently being presented. He was to repeat the word throughout the five seconds that he saw it and repeat it at a constant rate, about once every second.

After the subject was read the set of instructions appropriate for
### TABLE IV

**Graeco-Latin Square Design Used in Experiment II**

**Order of Presentation**

<table>
<thead>
<tr>
<th>Arrangement</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>List</td>
<td>Condition</td>
<td>List</td>
<td>Condition</td>
<td>List</td>
</tr>
<tr>
<td>1</td>
<td>C</td>
<td>Overt-Recall</td>
<td>B</td>
<td>Overt-Recognition</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>Silent-Recall</td>
<td>C</td>
<td>Silent-Recognition</td>
</tr>
<tr>
<td>3</td>
<td>A</td>
<td>Silent-Recognition</td>
<td>D</td>
<td>Silent-Recall</td>
</tr>
<tr>
<td>4</td>
<td>D</td>
<td>Overt-Recognition</td>
<td>A</td>
<td>Overt-Recall</td>
</tr>
</tbody>
</table>
his first two lists, the first list was presented. The click of an electric timer paced the experimenter’s presentation. Immediately following the completion of first-list presentation, the subject was handed an answer sheet appropriate to the condition and list under which he was being tested. He was given two minutes to complete the answer sheet. Then, he was told that he would be presented another list and he was to rehearse this list in the same way as he rehearsed the first list. Following the test of the second list, the subject was read instructions appropriate for rehearsal of the third and fourth lists. The third and fourth lists were then presented and tested.

Results

The subject’s protocols were scored as in Experiment I. The means (number correct minus number incorrect) were 27.60, 28.60, 12.75, and 10.00 for the Silent-Recognition, Overt-Recognition, Silent-Recall and Overt-Recall conditions, respectively. Means and variances for the four conditions are contained in Table V. Again, the data were transformed using a square root transformation. Transformed means and variances for the four conditions are also contained in Table V.

Analysis of variance was used to examine the transformed data in Table V. Variables that were examined were Lists (A, B, C, and D), Graeco-Latin square Arrangements (1, 2, 3, and 4), Presentation Order (the first, second, third, or fourth list learned by the subject), and Conditions (recognition vs. recall, silent vs. overt rehearsal). The results of this analysis are contained in Table VI. The significant effect of presentation order indicates that subjects performed best on
## TABLE V

Means and Variances and Transformed Means and Variance for the Four Conditions in Experiment II

<table>
<thead>
<tr>
<th>Measure</th>
<th>Recognition</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Silent</td>
<td>Overt</td>
</tr>
<tr>
<td>( \bar{x} )</td>
<td>27.60</td>
<td>28.60</td>
</tr>
<tr>
<td>( s^2_x )</td>
<td>41.73</td>
<td>40.15</td>
</tr>
<tr>
<td>( \sqrt{x} )</td>
<td>5.22</td>
<td>5.31</td>
</tr>
<tr>
<td>( s^2_{\sqrt{x}} )</td>
<td>0.40</td>
<td>0.42</td>
</tr>
</tbody>
</table>
TABLE VI

ANOVA Source Table for Conditions in Experiment II

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>ms</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total</strong></td>
<td>79</td>
<td>112.04</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>BETWEEN SUBJECTS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arrangements (1)</td>
<td>3</td>
<td>3.60</td>
<td>1.20</td>
<td>1.05</td>
<td>n.s.</td>
</tr>
<tr>
<td><strong>ERROR</strong></td>
<td>16</td>
<td>18.23</td>
<td>1.14</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>WITHIN SUBJECTS</strong></td>
<td>60</td>
<td>90.22</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Presentation Order (2)</td>
<td>3</td>
<td>0.92</td>
<td>0.31</td>
<td>3.59</td>
<td>&lt;.025</td>
</tr>
<tr>
<td>1 X 2</td>
<td>9</td>
<td>86.16</td>
<td>86.16</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Lists</td>
<td>3</td>
<td>2.50</td>
<td>0.83</td>
<td>9.79</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Recognition vs. Recall</td>
<td>1</td>
<td>76.93</td>
<td>76.93</td>
<td>905.04</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Silent Recall vs. Overt Recall</td>
<td>1</td>
<td>1.58</td>
<td>1.58</td>
<td>18.57</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Silent Recognition vs. Overt Recognition</td>
<td>1</td>
<td>0.08</td>
<td>0.08</td>
<td>&lt;1</td>
<td>n.s.,</td>
</tr>
<tr>
<td>Residual</td>
<td>3</td>
<td>5.07</td>
<td>1.69</td>
<td>21.12</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>ERROR</strong></td>
<td>48</td>
<td>4.06</td>
<td>0.08</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
the first list presented; this finding probably indicates the effect of proactive inhibition between lists. The significant effect of lists indicates that Lists B and C were easier to learn than Lists A and D. The instruction to rehearse overtly resulted in a significant decrement in recall performance but no significant decrement in recognition performance. Thus, Experiment II has replicated Experiment I by showing an effect of organization in recall but not in recognition. (The significant residual term contains too few degrees of freedom for further analysis. However, it does indicate that a between-subjects, rather than a within-subjects design be used if the experiment is to be replicated.)

A second analysis examined recall as a function of serial position for the Overt-Recall and Silent-Recall conditions. Probability of recall as a function of four blocks of 10 serial positions is graphed in Figure 2.

Analysis of variance was used to examine the data shown in Figure 2. The dependent measure was the number recalled at each of the four blocks of input positions for each subject. The results of the analysis are contained in Table VII. The significant effect of trials indicates the presence of a serial position effect. The significant Silent-Recall vs. Overt-Recall X Serial Position interaction indicates that the better recall of the Silent-Recall condition was limited to the primacy and middle portions of the serial position curve. This finding is in accord with the recall data reported by Fishler et al. (1970).

General Discussion

The data from the two experiments clearly suggest that the effect
FIGURE 2

Probability of Recall as a Function of Serial Position for Silent-Recall and Overt-Recall Conditions in Experiment II.
TABLE VII

ANOVA Source Table for Silent-Recall and Overt-Recall Conditions as a Function of Serial Position in Experiment II

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>ms</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOTAL</td>
<td>159</td>
<td>642.84</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>BETWEEN SUBJECTS</td>
<td>19</td>
<td>98.22</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>WITHIN SUBJECTS</td>
<td>140</td>
<td>544.63</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Silent Recall vs. Overt Recall (1)</td>
<td>1</td>
<td>18.91</td>
<td>18.91</td>
<td>7.66</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Serial Position (2)</td>
<td>3</td>
<td>170.02</td>
<td>56.67</td>
<td>22.96</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>1 X 2</td>
<td>3</td>
<td>27.48</td>
<td>9.16</td>
<td>3.71</td>
<td>&lt;.025</td>
</tr>
<tr>
<td>Error</td>
<td>133</td>
<td>328.23</td>
<td>2.47</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
of the organizational manipulations was on recall but not on recognition performance. The conclusion from this finding is that organization affects retrieval capacity but not storage capacity.

It is also clear that the organizational manipulations used in the present experiments, especially the imagery instruction, affected the form of the stored information; that is, the imagery instruction encouraged subjects to organize the items in groups of five and to use a visual code for these groups. However, the form of the stored information apparently had no effect on storage capacity as recognition scores did not vary as a function of the organizational manipulations. Thus, the evidence supports the contentions of Tulving (1963) and Kintsch (1970) and suggests that organization serves as a retrieval aid.

In addition to supporting the free recall theories of Tulving and Kintsch, the present data support Kintsch's (1970) two process theory of recognition and recall. Kintsch's two process theory includes the argument that there are qualitative, rather than quantitative, differences between recognition and recall, and it is counter to the notion that the superiority of recognition as compared to recall performance lies in the greater "strength" of items necessary for recall of those items. Data contrary to the "strength" notion would be showing the effect of an independent variable to differ as a function of whether recognition or recall tests are used. The present data, which showed that the organizational manipulations affected recall but not recognition performance, is an example of the type of data Kintsch (1970) cites as support for his two process theory.
One major difference between the two experiments involved the interaction between the recall conditions and serial position. In Experiment I, the imagery instruction resulted in greater recall across all blocks of serial position. In Experiment II, overt rehearsal resulted in depressed recall only in the primacy and middle portions of the serial position curve. This difference could imply that the imagery instruction affected the recency or short-term memory component of the input order. The appropriate test of this implication would involve showing improved performance in a Brown-Peterson (Brown, 1954; Peterson & Peterson, 1959) short-term memory paradigm when imagery instructions are used. Although such an experiment has not been conducted, Wickens and Engle (1970) have found that words rated high in imagery are recalled better in the Brown-Peterson paradigm than are words rated low in imagery.

The serial position curves found in Experiment II may be described in terms of opportunity for rehearsal. The elimination of the primacy effect for the overt rehearsal condition may be due to the loss of opportunity for extra rehearsal of early items. The failure to find differences between overt and silent rehearsal groups in recall of recent items may reflect no loss of opportunity for rehearsal of those items. Similar arguments have been advanced by Rundus and Atkinson (1970) and Fishler et al. (1970).

In conclusion, the two experiments reported in this chapter indicate that the effect of organization is to increase retrieval, but not storage, capacity. Implications of this finding in terms of a theory of free recall will be discussed in Chapter 7.
CHAPTER 4

CLUSTERING AND SUBJECTIVE ORGANIZATION

In Chapter 1, a major distinction between two types of organization, clustering and subjective organization, was made. Operationally, clustering and subjective organization differ in the types of items presented and the methods of measurement employed. For clustering, the items presented have an a\textit{ priori} structure, and clustering is measured in terms of the amount of conformity of the subject's recall protocols to that structure. For subjective organization, the items presented are unrelated, and subjective organization is measured in terms of the subject's tendency to repeat adjacencies from one trial to the next. Furthermore, subjective organization can be measured only in multitrial free recall, while clustering may be measured in single-trial free recall.

Similarities between the effects of experimental manipulations on clustering and subjective organization were also discussed in Chapter 1. Slowing presentation rate will result in higher recall and higher organization scores regardless of whether the materials presented have an a\textit{ priori} structure. Also, order of input can be manipulated so that increases in recall and organization scores can be observed for both categorized and unrelated lists. Cueing leads to increased recall from both lists of taxonomic categories and lists of subjective categories, and number of categories appears to have an effect on recall performance regardless of whether the categories are experimenter-determined or subject-determined. The data reported in Chapter 3 may also be interpreted as showing that an experimental manipulation has similar effects
on both categorized and unrelated lists. Instructing subjects to organize (Experiment I) or not to organize (Experiment II) lists of unrelated words affects recall and recognition performance in the same way that varying the strengths of category associates affects recall and recognition performance (Kintsch, 1968; Bruce & Fagan, 1970).

In Chapter 2, clustering and subjective organization were discussed in terms of Tulving's (1968) and Kintsch's (1970) theories of free recall. Tulving, apparently, does not make a distinction between the processes underlying clustering and subjective organization; however, he does not describe how properties of items are abstracted, or even what properties are abstracted, so that appropriate S-units are formed. Kintsch, on the other hand, appears to distinguish between clustering and subjective organization. Clustering is based on the similarities between the items on the list, while subjective organization may be idiosyncratic. However, Kintsch does not suggest different mechanisms underlying clustering and subjective organization.

Kintsch's assertion that similarity underlies clustering has been tested in an experiment reported by Kintsch, Miller, and Hogan (1970). In this experiment, some subjects were presented a list of 40 categorized words (five words from each of eight categories) and were asked to make similarity judgments on those words. Similarity judgments were made by having the subjects sort words into categories, and subjects were allowed to use anywhere between two and eight categories. The basic data were the number of subjects who sorted each pair of words into the same category; the more subjects who sorted each pair of words
into the same category, the more similar that pair was considered to be. Two self-paced sorting trials were used. Other subjects in the Kintsch et al. experiment were asked to free recall the same categorized list of words. The basic data from the free recall experiment were the number of times each pair was recalled adjacently. The subjects were given five free recall trials, and data from the first and fifth trials were analyzed. The free recall task was self-paced, and each trial was followed by 15 seconds of counting backwards.

The data from both the sorting and free recall tasks were used to generate four proximity matrices: one matrix each for the first sort, second sort, first recall trial, and fifth recall trial. The proximity matrices contained, for each pair of items, the number of subjects who, in the case of sorting, sorted that pair together or, in the case of free recall, recalled that pair adjacently. Each of the four proximity matrices was subjected to cluster analysis (BCTry: Tryon & Bailey, 1970). Marked similarity was found between the cluster structures produced from the sorting matrices, and both sorting and recall matrices were highly similar to the a priori structure of the list, thus supporting Kintsch's (1970) assertion that order of output in the free recall of a categorized list reflects similarity.

In the present chapter, an attempt was made to determine whether Kintsch's notion of similarities could be extended to a list of unrelated words. If order of output from a list of unrelated words can be predicted by similarity judgments, it would be additional evidence that the same processes underlie both clustering and subjective organization. In this
dissertation, similarity will be regarded as a measure of proximity from tasks in which subjects are asked to make "similarity" judgments.

Experiment III

Experiment III was designed to determine whether it is worthwhile to pursue the notion of similarity as a predictor of output order in free recall when the list of items does not have an a priori structure. In Experiment III, a nine-item list of "unrelated" words served as stimuli. Some subjects made similarity judgments on the nine words, and other subjects were given a multitrial free recall task with the same nine words serving as stimuli. Because of the short length of the list, a measure of similarity which yields more data per subject than does the card sorting task used by Kintsch et al. was employed. The measure was that of triads judgments (e.g., Wexler, 1970). For triads judgments, the subject is presented all possible sets of three of the stimulus words. The subject's task is, for each set, to select the word which he thinks is least similar to the other two words in that set.

Method

Stimuli and materials. The stimuli were nine words chosen randomly from Lists 3 and 4 of Tulving's (1964) lists of unrelated words, with the restriction that the six most infrequent words in each list not be chosen. For the triads task, all 84 possible sets of three words were used. The order of words within each set was randomly determined, and the entire set of 84 triads was randomly arranged for presentation. This arrangement of triads appeared in the answer booklets, with 24 triads on each of three pages and the remaining 12 triads on
For the free recall task, the nine words were mounted as slides for presentation via a carousel projector. Answer booklets contained 12 pages with nine blank lines per page.

**Design.** The design was correlational in nature. The basic data for both tasks were proximity matrices. The proximity matrix for the triads task contained the number of times, collapsed over subjects, each pair of words was chosen as more similar than the third word in the triad. The proximity matrix for the free recall task contained the number of times, collapsed over subjects and trials, each pair of words was recalled adjacently. The measure of the relationship between the two matrices was a Pearson product-moment correlation coefficient.

**Subjects.** For the triads task, the subjects were 17 volunteers from introductory psychology classes. The subjects were tested in groups of 1-3. For the free recall task, the subjects were 17 volunteers from undergraduate education classes. These subjects were tested in one group session.

**Procedure.** For the triads task, each subject was given an answer sheet and was told that, on each line of that sheet, there were three words. His task was to cross out the word on each line that he thought was least similar to the other two words on that line. The meaning of similarity was not defined. The task was self-paced.

For the free recall task, the subjects were read standard multi-trial free recall instructions which emphasized that the words could be recalled in any order. Words were presented at a one second rate, and there were 12 trials of free recall learning. There was a one minute
interval between trials during which the subjects were to recall as many words as they could. During this interval, the experimenter shuffled the slides in the slide tray to determine order of presentation for the next trial.

Results

Mean recall for the first trial was 6.56 items, and mean recall for the final trial was 9.00 items. The mean number of items recalled per trial ranged, across subjects, from 7.57 to 8.92 items, with an average of 8.42 items.

Proximity matrices for both tasks contained 36 cells, i.e. one cell for each possible pair of items. For the triads task, each pair was presented in seven triads. As there were 17 subjects in the triads task, the maximum possible cell entry was 119. For the free recall task, there were 12 trials and 17 subjects. The maximum possible cell entry, then, was 204.

The correlation between the proximity matrices for the free recall and triads tasks was \( r = .70 \). This correlation was highly significant, \( p < .001 \). However, the relationship between order of output and the triads judgments is underestimated. First, the data from the recall task included early trials in which short-term memory and input order are important determinants of output order. That is, the length of the list was little beyond the memory span, and a good portion of early-trial recall may have reflected the order of input. As all subjects had the same presentation orders, the effect of input order was not controlled. Second, although the instructions did not define similarity, the data
indicated that the triads subjects were rating mostly on similarity of
"meaning." However, the recall protocols indicated that other types of
similarity were used by the free recall subjects. For example, xylem
and zenith were the only two words on the list with the same initial
sound, and this pair was ranked second in frequency of adjacent recall.
However, xylem-zenith was the fifteenth ranked pair in frequency of
judged similarity. Elimination of xylem-zenith from the analysis raised
the correlation from .70 to .76.

Experiment IV

Experiment IV was designed as a replication and extension of Exper­
iment III. Experiment IV differed from Experiment III in that a longer
list of words was used and similarity judgments were obtained from a
card sorting task. For the card sorting, the basic data were, for each
pair of items, the number of subjects who placed that pair in the same
category. For the free recall task, the basic data were, for each pair,
the number of times that pair was recalled adjacently on the final three
trials. As in the Kintsch et al. (1970) study, proximity matrices were
generated from which cluster structures were derived. Descriptions of
the cluster analysis programme used and a method for comparing cluster
structures follow.

Cluster analysis. The cluster analysis programme used was HGroup
(Veldman, 1967). The form of the data for HGroup is an ordered vector
for each item. For example, in the present analysis, the vector for a
given word is the number of times it was sorted together (in the case of
card sorting) or recalled adjacently (in the case of free recall) with
every other word. The method of HGroup begins by treating each word as a category. At each step, the number of categories is reduced by one. For the first reduction, the two words which are closest, in terms of the minimum of the squared differences between their vectors, are combined. Then, the items which have been combined are given a vector which is the mean of the two vectors, and the process recycles. At the end of the method, there are two categories.

HGroup also yields selection values for optimal number of categories. These selection values, although not explicitly explained by Veldman, reflect the cost, in terms of squared deviations, of making the next reduction in the number of categories. The selection values are read so that the larger the selection value at a particular number of categories, the more reliable that category structure is.

Cluster comparison. HGroup will produce category structures from both the card sorting and free recall data. Then, it is necessary to compare the two cluster structures in order to determine if they are related. Schwartz and Humphreys (in preparation) have proposed a method for this comparison. Basically, two variables are involved in the analysis: (1) the number of specific pairs of items which are contained in both cluster structures; and (2) the number of specific pairs of items contained in the first structure but not in the second plus the number of specific pairs of items contained in the second structure but not in the first.

The analysis results in a chi-squared statistic with one degree of freedom. The formula for the chi-squared statistic involves the
deviation of the observed from the expected values of (1) and (2) described in the previous paragraph. Again, Schwartz and Humphreys should be referred to for the specific formula. As the analysis results in a chi-squared value, the contingency coefficient (see, e.g., Siegal, 1956) may be used to estimate the magnitude of the relationship.

Extension of Experiment III. Besides replicating Experiment III with a longer list of items and a different measure of similarity, Experiment IV was also designed to extend the conclusions of Experiment III. This extension was an attempt to predict recall performance as a function of the amount of conformity with the cluster structure produced by the card sorting task. That is, if the list of items were treated as if it had an a priori structure and organization were measured in terms of conformity to that structure, could the same relationships between organization and recall observed by Tulving (1962b) and Shapiro and Bell (1970), who used subjective organization scores, be demonstrated? The relationships observed by Tulving were (1) that increases in organization scores predicted increases in recall over trials and (2) that differences between subjects in organization scores predicted differences in recall. The relationships observed by Shapiro and Bell (1970) were that organization and recall increased over trials for high organizers, but recall increased over trials, without increases in organization, for low organizers. Simulations of the Tulving (1962b) and Shapiro and Bell (1970) results would provide additional evidence that the same processes underlie both clustering and subjective organization.1

1. The term simulation is used because the differences in measures of organization make the use of replication inappropriate.
Method

Stimuli and materials. The stimuli were 40 words chosen randomly from Tulving's (1964; Lists 1-4) lists of unrelated words, with the restriction that the six most infrequent words in each list not be chosen. For use in the card sorting task, each word was typed on a 3 x 5 inch card. Three sets of 40 cards each were prepared. Answer sheets for subjects in the card sorting task contained eight blank columns.

For the free recall task, the 40 words were mounted as slides. Also, ten subtraction problems using three-digit numbers were chosen, and these problems were mounted as slides. Answer booklets contained ten pages, with a space for the answer to the subtraction problem and 40 blank lines for the recall of words on each page. A carousel projector was used for presentation.

Design. The design contained a number of components, all of which were correlational in nature. The first of these components involved the replication of Experiment III. Proximity matrices were generated from both the sorting task and the free recall task. The basic data for these matrices were the number of times each pair of words was sorted into the same category or recalled adjacently for the sorting and recall tasks, respectively.

The second component involved the simulation of Tulving's (1962b) study of subjective organization in free recall when organization was measured as conformity to the categories produced by the cluster analysis of the sorting data. For each subject, a z-score was determined by treating each subject's ten trials as a single protocol. Then, clustering
and recall performance as a function of trials, collapsed over subjects, were compared. A comparison of $z$-scores and recall performance as a function of subjects, collapsed over trials, was also made.

The third component involved a simulation of Shapiro and Bell's (1970) study of subjective organization in free recall when organization was measured as conformity to the categories produced by the cluster analysis of the sorting data. On the basis of total $z$-scores, the subjects were divided into groups of high, middle, and low clusterers, and comparisons of $z$-scores and recall performance as a function of trials were made among these groups.

Subjects. For the sorting task, 40 volunteers from introductory psychology classes served as subjects. They were seen in groups of 1-3. For the free recall task, the subjects were 47 introductory psychology students. They were seen in one group session, and the experiment was run during their regularly-scheduled lecture period.

Procedure. For the sorting task, each subject was given a set of cards on which the 40 words were typed. The subject was told to go through the cards, one at a time and at his own pace, so that he could get a "feel" for the types of words that were on the cards. Then, he was to go through the deck again, this time sorting the words into categories. He was told to sort according to the meanings of the words, and he was to use anywhere between two and eight categories. The subject was allowed to change words or categories until he felt that he had made the best sort that he could. Then, he was to write down all the words in each category on his answer sheet. He was to use the columns to
represent categories. The subject was also told: (1) that he would not be tested on the words; (2) that there is no preconceived "correct" sort; and (3) that there is no preconceived "correct" number of categories. The order of the cards was randomized before the cards were given to the subject.

For the free recall task, the subjects were read standard multi-trial free recall instructions which emphasized that the words could be recalled in any order. They were told that, on each trial, they would be presented 40 words, one at a time, and that the last word would be followed by a subtraction problem. They were to answer the subtraction problem, and they could use the margin of the answer sheet to perform the arithmetic. After they recorded the answer to the subtraction problem, they were to write down as many words from the list as they could remember. The words were presented at a one second rate, and there was a two minute interval between trials. During this interval, the subjects (1) answered the subtraction problem, (2) recalled the words, and (3) turned the pages of their answer booklets, and the experimenter shuffled the slides in the slide tray to determine the order of presentation for the next trial. There were ten free recall learning trials.

Results

Cluster analysis. The mean and median numbers of categories used by the subjects in the card sorting task were 6.73 and 6.29, respectively. For each word, the number of subjects who sorted it with every other word formed the vector which was the input for the cluster analysis programme. For the entry which indicated the number of subjects who sorted a word
with itself, the value 40, which was the number of subjects in the card sorting task, was used. The data were analyzed both with and without the HGroup option to standardize the column scores. Identical solutions were obtained from both analyses, and the analysis with the standardization option is reported in this section. The selection values for the standardized solutions as a function of the number of categories are contained in Table VIII.

The selection values indicate that the most reliable solution with a reasonable number of categories is the five category solution, which has the highest selection value of any solution less than 24 categories. The category structure for the five category solution appears in Table IX.

For the free recall task, the data from the last three free recall trials were used. Mean recall scores for Trials 8, 9, and 10 were 27.51, 29.21, and 29.56, respectively. The last three trials were used because it was desired to obtain not only the most stable data but also sufficient data for the analysis. The data from four subjects were discarded from this and subsequent analyses because these subjects recalled in alphabetical order. Although recall in alphabetical order is one form of organization (Tulving, 1962a), it does not correspond with the card sorting subjects' instructions to sort according to similarity of meaning. In a similar manner, Kintsch et al. discarded data of subjects who recalled strictly according to input order.

For each word, the number of times it was recalled adjacently to every other word, collapsed over subjects and the final three trials, formed the vector which was the input for the cluster analysis programme.
TABLE VIII
Selection Values as a Function of Number of Categories for the Cluster Analysis of the Sorting and Free Recall Data of Experiment IV

<table>
<thead>
<tr>
<th>Number of Categories</th>
<th>Sorting</th>
<th>Free Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>39</td>
<td>14.97</td>
<td>9.22</td>
</tr>
<tr>
<td>38</td>
<td>15.98</td>
<td>259.78</td>
</tr>
<tr>
<td>37</td>
<td>9.94</td>
<td>15.77</td>
</tr>
<tr>
<td>36</td>
<td>9.91</td>
<td>28.13</td>
</tr>
<tr>
<td>35</td>
<td>6.04</td>
<td>1.82</td>
</tr>
<tr>
<td>34</td>
<td>0.00</td>
<td>1.87</td>
</tr>
<tr>
<td>33</td>
<td>24.05</td>
<td>10.07</td>
</tr>
<tr>
<td>32</td>
<td>0.30</td>
<td>1.30</td>
</tr>
<tr>
<td>31</td>
<td>39.95</td>
<td>2.10</td>
</tr>
<tr>
<td>30</td>
<td>1.52</td>
<td>3.91</td>
</tr>
<tr>
<td>29</td>
<td>8.64</td>
<td>0.80</td>
</tr>
<tr>
<td>28</td>
<td>0.54</td>
<td>2.16</td>
</tr>
<tr>
<td>27</td>
<td>0.63</td>
<td>0.55</td>
</tr>
<tr>
<td>26</td>
<td>2.12</td>
<td>0.04</td>
</tr>
<tr>
<td>25</td>
<td>8.92</td>
<td>1.65</td>
</tr>
<tr>
<td>24</td>
<td>8.27</td>
<td>1.01</td>
</tr>
<tr>
<td>23</td>
<td>0.02</td>
<td>0.17</td>
</tr>
<tr>
<td>22</td>
<td>1.29</td>
<td>0.40</td>
</tr>
<tr>
<td>21</td>
<td>3.88</td>
<td>0.53</td>
</tr>
<tr>
<td>20</td>
<td>1.53</td>
<td>0.47</td>
</tr>
<tr>
<td>19</td>
<td>1.19</td>
<td>0.18</td>
</tr>
<tr>
<td>18</td>
<td>1.08</td>
<td>0.07</td>
</tr>
<tr>
<td>17</td>
<td>2.54</td>
<td>0.72</td>
</tr>
<tr>
<td>16</td>
<td>1.35</td>
<td>0.47</td>
</tr>
<tr>
<td>15</td>
<td>1.65</td>
<td>0.55</td>
</tr>
<tr>
<td>14</td>
<td>1.07</td>
<td>0.21</td>
</tr>
<tr>
<td>13</td>
<td>1.18</td>
<td>0.57</td>
</tr>
<tr>
<td>12</td>
<td>0.73</td>
<td>0.46</td>
</tr>
<tr>
<td>11</td>
<td>0.92</td>
<td>0.50</td>
</tr>
<tr>
<td>10</td>
<td>1.78</td>
<td>1.30</td>
</tr>
<tr>
<td>9</td>
<td>0.02</td>
<td>0.57</td>
</tr>
<tr>
<td>8</td>
<td>0.94</td>
<td>0.04</td>
</tr>
<tr>
<td>7</td>
<td>1.84</td>
<td>0.43</td>
</tr>
<tr>
<td>6</td>
<td>3.95</td>
<td>0.51</td>
</tr>
<tr>
<td>5</td>
<td>7.10</td>
<td>0.27</td>
</tr>
<tr>
<td>4</td>
<td>1.05</td>
<td>0.12</td>
</tr>
<tr>
<td>3</td>
<td>4.76</td>
<td>0.28</td>
</tr>
<tr>
<td>2</td>
<td>-2.00</td>
<td>-2.00</td>
</tr>
</tbody>
</table>
# TABLE IX

Category Structures for the Sorting and the Free Recall Data in Experiment IV

<table>
<thead>
<tr>
<th>Word</th>
<th>Sorting</th>
<th>Free Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Action</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Effort</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Impact</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Answer</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Question</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Finding</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Rumor</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Treason</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Waiver</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Express</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>Accent</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>Bridle</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Flower</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Gable</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Noodle</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Ether</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>Water</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>Vulture</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>Letter</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>Hermit</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Miser</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Gambler</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>Orphan</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>Voter</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>Despot</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>Walker</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>Novice</td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>Buyer</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>Cherub</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>Legion</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Union</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Maxim</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>Zenith</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>Office</td>
<td>4</td>
<td>10</td>
</tr>
<tr>
<td>Quarter</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>Centre</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>Island</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Ocean</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Jungle</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Valley</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>
For the entry which indicated the number of times a word was recalled adjacently to itself, the value 80 was used. Again the standardization option was employed. The selection values for solutions as a function of a number of categories are contained in Table VIII.

The selection values indicate that the most reliable solution with a reasonable number of categories is the ten category solution, which has the highest value of any solution less than 25 categories. The category structure for the ten category solution appears in Table IX.

The number of specific pairs of items shared in common by the two category structures was 36. The number of pairs contained in the card sorting structure but not in the free recall structure plus the number of pairs contained in the free recall structure but not in the card sorting structure was 183. Under the hypothesis that there is no relationship between the two structures, the expected number of pairs shared in common is 19.87 and the expected number unique to either structure is 163.13. Deviations of the observed from the expected leads to a significant chi-squared statistic, $\chi^2 (1) = 14.68, p < .001$. Thus, there is a relationship between the two cluster structures.

Tulving (1962b)-type analysis. Conformity to the category structure produced by the cluster analysis of the sorting data (see Table IX) was analyzed according to the multiple category runs test (Frankel & Cole, 1971). A $z$-score was obtained for each subject for each trial. The mean $z$-score per trial is illustrated in Figure 3. As with 43 subjects, a mean $z$-score of 0.39 is needed for significance at the .01 level, clustering on all trials except the first was statistically significant.

Mean recall per trial is also illustrated in Figure 3. A
FIGURE 3

Recall and Organization Scores as a Function of Trials
for All Subjects in Experiment IV

![Line graph showing recall and organization scores as a function of trials for all subjects in Experiment IV. The x-axis represents trial numbers from 1 to 10, and the y-axis represents recall and z-score values. The graph shows an increasing trend in both recall and z-score as the number of trials increases.]]
correlation between mean recall and mean z-scores as a function of trials was $r = .94, p < .001$. This correlation is very similar to that of .96 reported by Tulving (1962b) who measured organization by subjective organization scores. Thus, the increase in recall of a list of unrelated words can be predicted by increases in conformity to a category structure derived from similarity data.

To calculate a total clustering score for a subject, his ten recall protocols were treated as a single protocol. From this protocol, a z-score was obtained. The mean of these z-scores was 4.65 which is statistically significant. Total recall scores were also obtained for each subject. The correlation between total recall and total z-scores as a function of subjects was statistically significant, $r = .49, p < .002$. This correlation is similar to that reported by Tulving (1962b) who found a correlation of .62 between total recall and total subjective organization scores. Thus, differences among subjects in the recall of unrelated words can be predicted from their conformity to a category structure derived from similarity data.

Shapiro & Bell (1970)-type analysis. On the basis of the total clustering scores, subjects were divided into three groups. The 13 subjects with the highest z-scores were designated as high organizers, the 13 subjects with the middle z-scores as middle organizers, and the 13 subjects with the lowest z-scores as low organizers. Mean z-scores for the high, middle, and low organizers were 5.28, 2.37, and 0.38, respectively. The z-score for the low organizers indicates that they did not cluster more than would be expected by chance, $z = 1.39, p > .05$. High
organizers recalled a mean of 253.31 words, middle organizers recalled a mean of 229.92 words, and low organizers recalled a mean of 194.46 words.

Recall and \( z \)-scores as a function of trials were examined for the three groups. Mean recall per trial for the three groups is shown in the upper part of Figure 4, and mean \( z \)-scores as a function of trials for the three groups is shown in the lower part of Figure 4.

Correlations between recall and \( z \)-scores as a function of trials were obtained for the three groups. The correlations between the recall curves and the \( z \)-score curves for the high and middle organizers were \( r = .96 \) and \( r = .91 \), respectively. The correlation between the two curves for the low organizers was \( r = .07 \), which is not statistically significant. Thus, recall increased over trials for high, middle, and low organizers and organization increased over trials for high and middle organizers. However, the increase in recall over trials for low organizers was not accompanied by a corresponding increase in their organization scores. These results are similar to those of Shapiro and Bell (1970), who measured subjective organization rather than conformity to a category structure.

Experiment V

Experiment IV replicated the most important results from correlational studies of the subjective organization and recall of "unrelated" words by treating a list of unrelated words as a categorized list, with categories derived from similarity judgments. However, one aspect of clustering according to these categories, as compared to clustering
FIGURE 4

Upper: Recall as a Function of Trials for High (Ho), Middle (Mo), and Low (Lo) Organizers in Experiment IV.

Lower: Organization as a Function of Trials for High (Ho), Middle (Mo), and Low (Lo), Organizers in Experiment IV.
according to categories taken from category norms, was not reproduced in Experiment IV. That is, significant clustering was not observed in the first trial of Experiment IV, while it is usually observed in the single-trial free recall of lists taken from norms. Experiment V was an attempt to demonstrate that significant clustering, defined as conformity to the categories derived from the cluster analysis of the sorting data in Experiment IV, could be observed in a single-trial free recall.

The failure to observe significant first-trial clustering in Experiment IV could have been due to the small mean number of items recalled, 9.53. This mean is little beyond the span of immediate memory, and, thus, first-trial recall might have reflected presentation order. As the presentation order did not vary across subjects, any effect of first-trial clustering may have been obscured. In line with the preceding argument, Experiment V differed from Experiment IV in two ways. First, the presentation rate was slowed from one second to five seconds. Second, the order of presentation was randomized for each subject in the experiment.

Method

Stimuli and materials. The list of 40 words used in Experiment IV served as stimuli. For presentation, a set of 40 3 x 5 inch cards which contained those words was used. An answer sheet containing 40 blank lines was prepared for each subject.

Design. No experimental manipulation was used. All subjects were presented the same list of words. It was desired to obtain a z-score, which was the measure of category clustering, for each subject, and to
determine whether the mean $z$-score indicated that subjects were clustering beyond the chance level.

**Subjects.** The subjects were ten volunteers from introductory psychology classes. They were tested individually.

**Procedure.** The subjects were read standard single-trial free recall instructions which emphasized that the words could be recalled in any order. The words were then presented at a five second rate. The click of an electric timer paced the experimenter's presentation of the words. Following the presentation of the last word, the experimenter handed the subject an answer sheet on which the subject was to write down as many words as he could remember. The subject was given two minutes to recall the words. Order of presentation was randomized for each subject.

**Results**

Mean recall was 14.50 words. This was greater than the mean of 9.53 observed on the first trial of Experiment IV. The mean $z$-score, 1.10, was also greater than that observed on the first trial of Experiment IV, 0.09. The mean $z$-score indicated significant clustering according to the category structure derived from the sorting task in Experiment IV, $z = 3.48$, $p < .001$. Thus, significant clustering in a list of unrelated words which has been categorized according to similarity judgments has been observed in single-trial free recall.

**Discussion**

The experiments reported in Chapter 4 have served to reduce the distinction between the clustering of items from a categorized list and the subjective organization of items from an unrelated list. First,
Experiments III and IV have shown that some agreement can be reached by subjects who judge the similarity of items in an "unrelated" list, and that these similarity judgments can predict order of output when other subjects free recall these items. Thus, Experiments III and IV have replicated Kintsch et al.'s (1970) findings of a correlation between similarity judgments and output order, except that the lists used by Kintsch et al. had an a priori structure taken from category norms while the lists used in Experiments III and IV did not have such a structure.

Second, Experiment IV showed that the amount of clustering according to the category structure derived from similarity judgments of a list of "unrelated" words was a predictor of free recall performance. Mean recall and mean clustering scores as a function of trials were found to be highly correlated. A substantial correlation was also found between recall scores and clustering scores as a function of subjects. Thus, amount of organization predicted both increases in recall over trials and differences among subjects in total recall, simulating the results of Tulving's (1962b) experiment. However, Tulving, who also used unrelated words, measured organization according to subjective organization scores while Experiments III and IV measured organization according to clustering scores. The difference is that Tulving's measure tapped the amount of consistency in a subject's output order from one trial to the next, while the measure used in Experiment IV tapped the amount of conformity of a subject's output order with a category structure derived from similarity judgments.

Third, differences between high, middle, and low organizers were
found in Experiment V. For high and middle organizers, there were systematic increases in recall but no systematic increases in clustering scores over trials. Thus, Experiment IV has simulated the results of Shapiro and Bell (1970) who found that the recall of low organizers increased over trials without corresponding increases in organization scores. The differences between Experiment IV and the Shapiro and Bell study are the same as the differences between Experiment IV and the Tulving (1962b) study.

Fourth, Experiment IV found that the organization of unrelated words could be observed in a single-trial free recall. As one of the operational differences between the subjective organization of unrelated words and the clustering of words which have an a priori structure is that the former can only be observed in multitrial free recall while the latter may be observed in single-trial free recall, the results of Experiment V reduce the distinction between clustering and subjective organization.

Although the results of the experiments reported in this chapter correspond with the results of previous studies, some differences are apparent. Particularly, the correspondence between the cluster structures derived from recall and sorting data reported by Kintsch et al. is much greater than that found in Experiment IV. Also, middle organizers in Experiment IV showed systematic increases in organization over trials, but middle organizers in the Shapiro and Bell study did not show such increases.

There are a number of major differences between the Kintsch et al. study and Experiment IV which may have led to the greater correspondence
between the cluster structures derived from sorting and those derived from recall observed by Kintsch et al. First, Kintsch et al. probably observed greater recall and thus, more stable output orders, than those observed in Experiment IV. Kintsch et al. used a self-paced presentation rate and Experiment IV used a one second presentation rate. Also, the obvious category structure of the lists used by Kintsch et al. probably led to greater recall than from the unrelated words used in Experiment IV. If one chooses to view the list used in Experiment IV as composed of the categories derived from the sorting data, then the words used by Kintsch et al. should be thought of as more obvious category members than the words used in Experiment IV. The use of more obvious categories may have led to a greater agreement between card sorting subjects in the Kintsch et al. study. That is, the categories used in the Kintsch et al. study may have been mutually exclusive, such that a word sorted into one category would almost never be sorted into any other category. On the other hand, an examination of the category structure found in Experiment IV (see Table IX) suggests a number of words which may have been sorted into a category other than that in which they appear. Thus, the recall protocols could have reflected, to some extent, these other possible categorizations.

A third reason for the superior correspondence found in Kintsch et al's study may have been that the effects of input order were controlled in their study, but not in Experiment IV. That is, for each subject on each trial, Kintsch et al. randomized the presentation order; in Experiment IV, on each trial all of the subjects received the same presentation
order. Thus, the category structure derived from the free recall data in Experiment IV may have reflected input order more than that derived from the Kintsch et al. data.

A second discrepancy between Experiment IV and a previously-published experiment was that middle organizers showed systematic increases in organization over trials in Experiment IV, but middle organizers may not have shown such systematic increases in Shapiro and Bell's (1970) study. Thus, the present findings question Shapiro and Bell's conclusion that middle organizers show increases in recall over trials without corresponding increases in organization scores. The major difference between Shapiro and Bell's study and Experiment IV was that organization was measured by subjective organization scores in the former, but organization was measured by clustering scores in the latter. Clustering scores may be more sensitive than subjective organization scores because, given the recall of a particular item on trial \(n\), an increase in subjective organization requires the adjacent recall of the specific item which was recalled adjacently on trial \(n-1\), while an increase in clustering requires the adjacent recall of any item that belongs to the same category as the recalled item (see Chapter 1).

If the argument that a more sensitive measure of organization was used in Experiment IV than in the Shapiro and Bell study is accepted, an implication about experiments investigating the necessity of increases in organization for increases in recall performance is suggested. That is, if the sensitivity of the organization measure used determines whether increases in organization scores are found, studies in which increases in
recall performance without corresponding increases in organization scores have been observed may have failed to use a sufficiently sensitive measure of organization. By extension, then, the clustering measure used in Experiment IV may not have been sensitive enough to detect increases in organization for low organizers. A similar argument concerning the sensitivity of the organization measure has been made both by Handler (1967) and in the first chapter of the present paper.

In conclusion, the experiments reported in the present chapter have reflected striking parallels between the subjective organization of a list of unrelated words and the clustering of a list of categorized words. Although the present experiments have not proved that similar processes underlie both clustering and subjective organization, they strongly suggest a conclusion involving similar processes. Such a conclusion would not necessitate the exclusion of idiosyncratic organization in the recall of unrelated words, but idiosyncratic organization has not been excluded in the recall of categorized words. For present purposes, the clustering of categorized words and the subjective organization of unrelated words will not be treated as qualitatively different, and the informal model presented in Chapter 7 will not describe different processes for the two types of organization.
CHAPTER 5
AN EXAMINATION OF ORDER OF OUTPUT

The research reported in this chapter was conducted for four purposes. The first of these purposes was to extend the findings reported in Chapter 4 and by Kintsch et al. (1970). The data reported in Chapter 4 indicated that similarity judgments can predict order of output from a list of unrelated words, and the data reported by Kintsch et al. indicated that similarity judgments can predict order of output from a list of categorized words. In the present chapter, two types of items different from those employed either in Experiments III, IV, and V or by Kintsch et al. were used, and order of output was compared with data from similarity judgments. In Experiment VI the items were all taken from what could be considered one category, and in Experiment VII the items were pairs of words, with both members of a pair belonging to the same category.

The second purpose of the experiments reported in this chapter was to compare order of output in free recall with data from similarity judgments as a function of the type of analysis used. The items used in Experiment VI had been stimuli in various similarity judgments tasks from which cluster structures and multidimensional structures were derived (Henley, 1969). In Experiment VI, order of output in free recall was used to generate proximity matrices. From these matrices, cluster structures and multidimensional structures were derived, and these structures were compared with those derived from the similarity judgments data.

A third purpose of the experiments was to provide a description of
order of output in free recall and to examine this description in terms of theories of free recall. That is, theories such as Tulving's (1968), Mandler's (1967), and Kintsch's (1970) describe a subject's internal structure which accounts for his order of output. Kintsch's theory even contains a mechanism for extracting information from long-term memory so that similar items are recalled adjacently. However, none of these theories is a process theory in terms of describing order of output; that is, although the theories could be modified so that they are able to describe which items are most frequently recalled together, they cannot fully predict the order of output in a given subject's recall protocols. The description of output order for some individual subjects in Experiment VI may help provide a link between output and internal structure.

A fourth purpose of the experiments in this chapter was to differentiate between the three types of internal structures of organization: clusters, hierarchical clusters, and multidimensional spaces. That is, certain results from the cluster analysis, multidimensional scaling, or examination of output order may be incompatible with one of the internal structures or may suggest a modification of one of the internal structures. Rather than review the results which would suggest possible modifications, the discussion of the implications for types of internal structures will be made in terms of the results of Experiment VI. Also, Experiment VII was designed to determine whether subjects will use hierarchical categories, and whether the formation of hierarchical categories can predict recall performance.
Experiment VI

Experiment VI was similar to Experiments III and IV in that data from similarity judgments tasks were compared with order of output from a free recall task. However, Experiment VI differed from Experiment III and Experiment IV in the following ways: (1) the items were drawn from what might be considered to be one category; (2) the similarity judgments were from a previously published study; (3) the comparison was made as a function of two types of analyses; and, (4) a different distance measure was used in examining order of output in free recall. The stimuli, the distance measure, and the types of analyses will be discussed separately.

Stimuli. The items were 30 mammal names which had been used in experiments reported by Henley (1969). Using data from various similarity judgments tasks, Henley derived cluster structures and multidimensional structures for these items. The exact items and a typical cluster structure are contained in Table XI. The typical multidimensional solution contained three dimensions. These dimensions might loosely be labeled size, ferocity, and human-ness (Henley, 1969). The extremes of the size dimension were mouse and elephant; the extremes of the ferocity dimension were cow and tiger; and the extremes of the human-ness dimension were pig and gorilla.

Distance measure. In Experiments III and IV the measure of distance in free recall protocols was the number of times each pair of items was recalled adjacently. In Experiment VI the measure used was one of relative distance (see, e.g., Buschke & Hinrichs, 1968). For any recalled item, there are two items which are recalled closest to it; these two
items are given a relative distance measure of one. The next two closest items to the recalled item are given a relative distance measure of two, and so on. In Experiment VI the relative distance between each recalled item and every other recalled item was determined for each subject's final recall protocol. The relative distance between each item and all other items was then averaged across subjects, and a proximity matrix was obtained. This proximity matrix was used as input for the cluster analysis programme, and the matrix was collapsed into lower diagonal form for the multidimensional scaling programme. As, so far as the present author is aware, relative distance has not been used in previous studies of order of output in free recall, the validity of the relative distance measure can only be described as a function of the results of Experiment VI. Relative distance was used because it provides more information from a single protocol than does number of adjacencies.

Analysis. The cluster analysis programme used was HGroup, which was described in Chapter 4. For a hierarchical cluster analysis, HGroup was also used. That is, although HGroup yields a selection value for the most reliable number of categories, the programme will continue to combine categories past the number of categories which yields the highest selection value. This combination results in a hierarchical cluster structure.

The multidimensional scaling programme used was SSA-1 which is contained in the Guttman-Lingoes package of scaling programmes. SSA-1 has been described by Guttman (1968), and an empirical comparison between SSA-1 and other multidimensional scaling programmes has been reported by
Lingoes and Roskam (1971). The results of the multidimensional scaling of the group data did not sufficiently resemble the results of Henley's (1969) analysis. Because the failure to find sufficient resemblance cannot be traced to one specific source, the multidimensional scaling will not be further discussed.

Method

Stimuli and materials. The stimuli were 30 mammal names taken from Henley (1969). Each of these was typed on a 3 X 5 inch card. Seven three-digit numbers were also chosen to be used as starting numbers for counting backwards following each presentation of the list. Answer sheets were prepared and contained seven pages with 30 blank lines per page.

Design. No experimental manipulation was used. Each subject was presented the same list of items for seven trials of presentation and recall. Relative distance measures were obtained from each subject's seventh recall trial, and these distances were averaged over subjects for use as input for the cluster analysis programme. Comparisons between the solutions obtained from the free recall protocols and Henley's data were made. Also, an examination of recall protocols for two subjects who recalled all 30 items on Trials 4-7 was made. The relative distance between each pair of items for each of the last four trials was obtained, and these relative distances were averaged over the four trials to form a proximity matrix for each subject.

Subjects. The subjects were 20 volunteers from introductory psychology classes. They were tested individually.

Procedure. The subject was told that he would be given a deck of
30 cards and that each card contained a mammal name. He was to go through the deck at his own pace, reading each card aloud, studying it, and then turning it face down. He was not allowed to go back to a card once he had turned it face down. The subject was told that after he turned over the last card, he would be read a three-digit number. He was to count backward by ones from this number for 15 seconds. After 15 seconds he would be told to stop counting and asked to write down on the first page of his answer sheet, in any order he desired, as many words from the list as he could remember. When he could think of no more words, he was to turn the page of his answer sheet. Then, the same procedure would be repeated, as it would for a total of seven trials. During the intervals in which the subject recalled, the experimenter shuffled the cards to determine order of presentation for the next trial.

Results

Performance. Mean recall per trial is shown in Figure 5. The mean recall of over 28 items on Trial 7 indicates that fairly complete data were available for the relative distance measure. The subject recalling the fewest items on Trial 7 recalled 25 items on that trial. The amount recalled by subjects ranged from 18.6 items per trial to 29.1 items per trial, over the seven trials.

The cluster structure derived from Henley's (1969) data, shown in Table XI, was used to treat the list as if it had an *a priori* structure. Clustering was measured for each subject on each trial according to conformity to that structure. The clustering measure used was the $z$-score (see Chapter 1). Mean $z$-scores per trial are shown in Figure 5. As can
FIGURE 5

Recall and Z-scores as a Function of Trials for Experiment VI
be seen from Figure 5, the $z$-scores indicate that clustering was significant at least at the .001 level on each trial. A correlation between the $z$-scores and recall scores was found to be highly significant, $r = .96$.

A total $z$-score was calculated for each subject by treating the seven trials as a single protocol. The range of the $z$-scores thus obtained was from 3.20 to 15.02, and the mean was 8.15. These $z$-scores indicate that all subjects were clustering at less than the .001 level. A correlation between the total $z$-scores and recall performance was positive but did not reach significance, $r = .36$, $p > .05$.

The conclusion from this section is that subjects clustered according to the categories derived from similarity judgments and that the amount of this clustering increased over trials along with increases in recall performance. The failure to find a significant correlation between recall performance and $z$-scores may be due to the self-paced presentation or the small number of subjects, rather than to the nature of the materials.

Group cluster analysis. Each item was described by a vector which contained the mean Trial 7 relative distances between that item and every other item. The value 1.0, the minimum of possible observed relative distances, was used to describe the diagonal entries indicating the relative distance between an item and itself. These vectors formed the proximity matrix which was the input for HGroup. As in Experiment IV, the HGroup option to standardize the column entries was used.

Table X contains the selection values for solutions as a function of the number of categories. These selections values indicate that the eight category solution was the most optimal solution containing fewer than 18 categories.
TABLE X

HGroup Selection Values for Group Data and Data for Two Individual Subjects as a Function of Number of Categories in Experiment VI

<table>
<thead>
<tr>
<th>Number of Categories</th>
<th>Group Data</th>
<th>Subject 1</th>
<th>Subject 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>29</td>
<td>56.51</td>
<td>6.41</td>
<td>6.30</td>
</tr>
<tr>
<td>28</td>
<td>4.22</td>
<td>2.77</td>
<td>27.27</td>
</tr>
<tr>
<td>27</td>
<td>3.12</td>
<td>11.59</td>
<td>7.70</td>
</tr>
<tr>
<td>26</td>
<td>1.81</td>
<td>1.00</td>
<td>6.75</td>
</tr>
<tr>
<td>25</td>
<td>16.10</td>
<td>6.56</td>
<td>0.63</td>
</tr>
<tr>
<td>24</td>
<td>2.13</td>
<td>4.19</td>
<td>2.96</td>
</tr>
<tr>
<td>23</td>
<td>2.94</td>
<td>10.64</td>
<td>1.07</td>
</tr>
<tr>
<td>22</td>
<td>2.24</td>
<td>0.31</td>
<td>0.07</td>
</tr>
<tr>
<td>21</td>
<td>2.71</td>
<td>3.10</td>
<td>0.48</td>
</tr>
<tr>
<td>20</td>
<td>0.38</td>
<td>1.65</td>
<td>7.06</td>
</tr>
<tr>
<td>19</td>
<td>0.36</td>
<td>0.36</td>
<td>5.18</td>
</tr>
<tr>
<td>18</td>
<td>23.77</td>
<td>0.78</td>
<td>9.16</td>
</tr>
<tr>
<td>17</td>
<td>0.86</td>
<td>2.61</td>
<td>4.66</td>
</tr>
<tr>
<td>16</td>
<td>0.41</td>
<td>9.08</td>
<td>10.80</td>
</tr>
<tr>
<td>15</td>
<td>2.38</td>
<td>13.33</td>
<td>3.20</td>
</tr>
<tr>
<td>14</td>
<td>3.59</td>
<td>1.28</td>
<td>1.62</td>
</tr>
<tr>
<td>13</td>
<td>3.61</td>
<td>4.73</td>
<td>3.29</td>
</tr>
<tr>
<td>12</td>
<td>1.05</td>
<td>0.89</td>
<td>0.20</td>
</tr>
<tr>
<td>11</td>
<td>1.66</td>
<td>2.09</td>
<td>1.64</td>
</tr>
<tr>
<td>10</td>
<td>0.63</td>
<td>7.56</td>
<td>1.01</td>
</tr>
<tr>
<td>9</td>
<td>0.11</td>
<td>1.43</td>
<td>3.02</td>
</tr>
<tr>
<td>8</td>
<td>5.29</td>
<td>1.79</td>
<td>7.77</td>
</tr>
<tr>
<td>7</td>
<td>1.35</td>
<td>2.79</td>
<td>0.80</td>
</tr>
<tr>
<td>6</td>
<td>3.13</td>
<td>0.00</td>
<td>2.33</td>
</tr>
<tr>
<td>5</td>
<td>0.06</td>
<td>0.20</td>
<td>0.95</td>
</tr>
<tr>
<td>4</td>
<td>1.25</td>
<td>7.36</td>
<td>6.39</td>
</tr>
<tr>
<td>3</td>
<td>0.44</td>
<td>3.53</td>
<td>5.37</td>
</tr>
<tr>
<td>2</td>
<td>-2.00</td>
<td>-2.00</td>
<td>-2.00</td>
</tr>
</tbody>
</table>
The category structure for the eight category solution is illustrated in Table XI. This category structure was compared with Henley's category structure using the Schwartz and Humphreys' (in preparation) method for category comparison. A chi-squared statistic indicated that the free recall category structure and Henley's category structure were significantly related $\chi^2(1) = 71.27$, $p < .001$. The associated contingency coefficient was $C = .46$. The conclusion from these statistics is that the cluster structures derived from similarity judgments data and from order of output in free recall show a high degree of resemblance.

The combination of categories into fewer than eight categories results in a hierarchical cluster structure. The hierarchical cluster structure for the group data is illustrated in Figure 6. Although Henley did not report a hierarchical arrangement of her categories, the hierarchical arrangement shown in Figure 6 appears to be intuitively pleasing as the nodes are fairly obvious. Possible names for nodes are illustrated in parentheses in Figure 6. For the most part, with the exception of the chipmunk-squirrel category, the items contained in higher-order nodes are fairly well described by those nodes. The conclusion is that a hierarchical cluster structure may provide a good description of order of output in free recall. However, this conclusion must be viewed with caution because the hierarchical arrangement may simply reflect the differing probabilities of items being contained in categories. That is, the observation of a sensible hierarchical structure may have occurred as a function of averaging relative distances across subjects. This possibility was more directly tested in Experiment VII which was designed to determine
<table>
<thead>
<tr>
<th>Stimulus</th>
<th>Henley</th>
<th>Group Data</th>
<th>Subject 1</th>
<th>Subject 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chimpanzee</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Gorilla</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Monkey</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Beaver</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Raccoon</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Rabbit</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Chipmunk</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Squirrel</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Bear</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Elephant</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Giraffe</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Leopard</td>
<td>3</td>
<td>5</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>Lion</td>
<td>3</td>
<td>5</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>Tiger</td>
<td>3</td>
<td>5</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>Mouse</td>
<td>4</td>
<td>6</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>Rat</td>
<td>4</td>
<td>6</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>Cat</td>
<td>4</td>
<td>6</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>Dog</td>
<td>4</td>
<td>6</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>Fox</td>
<td>4</td>
<td>7</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>Wolf</td>
<td>4</td>
<td>7</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>Antelope</td>
<td>5</td>
<td>4</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>Deer</td>
<td>5</td>
<td>2</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>Cow</td>
<td>5</td>
<td>8</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Pig</td>
<td>5</td>
<td>8</td>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td>Sheep</td>
<td>5</td>
<td>8</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>Gcat</td>
<td>5</td>
<td>8</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>Camel</td>
<td>5</td>
<td>4</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>Donkey</td>
<td>5</td>
<td>8</td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td>Horse</td>
<td>5</td>
<td>8</td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td>Zebra</td>
<td>5</td>
<td>4</td>
<td>8</td>
<td>2</td>
</tr>
</tbody>
</table>
FIGURE 6
Hierarchical Category Structure for the Group Data in Experiment VI
if subjects will form hierarchical cluster structures.

**Individual Cluster Analysis.** Two subjects recalled all 30 items on each of Trials 4-7. Relative distance measures were obtained for these two subjects on each of the four trials, and the measures were averaged across the four trials. Thus, a proximity matrix was obtained for each of the two subjects, and these proximity matrices served as input for the cluster analysis programme. As in previous analyses, the value of 1.0 was used for the diagonal entries, and the standardization option was employed.

Selection values for the two subjects as a function of number of categories are contained in Table X. For one subject, the selection values indicate that the ten category solution was the best solution containing fewer than 15 categories. The ten category solution for this subject is shown in Table XI. For the other subject, the selection values indicate that the eight category solution was the best solution containing fewer than 16 categories. The eight category solution for this subject is shown in Table XI. Chi-squared values were computed for the relationships between the cluster structure for the two subjects, the group data, and Henley's (1969) data; with each chi-squared value, associated contingency coefficients were obtained. The chi-squared values and contingency coefficients are reported in Table XII. All chi-squared values found in Table XII are significant at less than the .025 level. Thus, the cluster structures derived from similarity judgments data can, to some extent, predict the cluster structures derived from individual recall protocols.


TABLE XII

Chi-squared Values and Contingency Coefficients for Relationships between Cluster Structures for Henley's (1969) Data and Group Data and Two Individual Subjects in Experiment VI

<table>
<thead>
<tr>
<th>Structures</th>
<th>Chi-squared</th>
<th>Contingent Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Henley - Group</td>
<td>71.27</td>
<td>.46</td>
</tr>
<tr>
<td>Henley - Subject 1</td>
<td>62.89</td>
<td>.58</td>
</tr>
<tr>
<td>Henley - Subject 2</td>
<td>5.75</td>
<td>.20</td>
</tr>
<tr>
<td>Group - Subject 1</td>
<td>115.92</td>
<td>.75</td>
</tr>
<tr>
<td>Group - Subject 2</td>
<td>59.61</td>
<td>.61</td>
</tr>
<tr>
<td>Subject 1 - Subject 2</td>
<td>5.38</td>
<td>.24</td>
</tr>
</tbody>
</table>
The combination of categories into hierarchical categories for the two subjects is illustrated in Figures 7 and 8. The hierarchical arrangement shown in Figure 7 seems to be sensible, and it appears to correspond, to some extent, with the hierarchical arrangement for the group data. The hierarchical arrangement shown in Figure 8 seems to be less sensible and has less correspondence to the group data. The conclusion is that the individual data can be described by a model of hierarchically-arranged categories, although the particular hierarchies may vary across subjects.

Order of output. Examination of the recall protocols for the two subjects indicated that both subjects were recalling in approximately the same order on the four trials that were analyzed. That recall was in a fairly constant order over the last four trials for these subjects has an important implication about a model used to describe free recall. That is, a model would require a mechanism which would retrieve items or groups of items from memory in a fixed order. Such a mechanism would be compatible with a hierarchical arrangement of items or groups of items.

Experiment VII

The data reported in Experiment VI point to the need for further investigation of hierarchically-arranged categories. In Experiment VII, three questions about hierarchical clustering were asked. First, will subjects form hierarchical clusters? Second, will increases in hierarchical clustering vary with increases in recall as a function of trials? Third, can the arrangement of categories into higher-order categories be predicted by similarity judgments data?
FIGURE 7
Hierarchical Category Structure for
Individual Subject 1 in
Experiment VI
Hierarchical Category Structure for Individual Subject 2 in Experiment VI

FIGURE 8
Method

Stimuli and materials. The stimuli were 12 pairs of items; both members of a pair were high associates to the same category. The categories were chosen randomly from the Battig and Montague (1969) norms, and there were no obvious relationships among the categories.

For presentation in the free recall task, two decks of 12 3 X 5 inch cards were used. Each deck contained the 12 pairs with one pair on each card. A given card contained one member of the pair typed above the other member of that pair. The difference between the decks was that the top member of the pair was interchanged between decks. That is, if Deck A contained red typed above blue, Deck B contained blue typed above red. Answer booklets for the recall task contained seven pages, and there were 12 pairs of blank lines on each page. Seven three-digit numbers were chosen to be used as starting numbers for counting backwards following each presentation of the list.

For use in the triads task, all 220 possible combinations of three of the pairs were typed on 3 X 5 inch cards. The pairs were typed from left to right on the cards, and one word in each pair was typed above the other word in that pair. Each pair was arbitrarily assigned a number from 1-12, and the number assigned to a pair appeared above that pair. The order of the pairs was randomly arranged from left to right on each card, and the member of a pair which appeared on top was randomly determined for each pair on a card. The cards were numbered from 1-220, and the number of a card appeared in its upper right-hand corner. Answer sheets for the triads task contained 220 pairs of blank lines, and the
pairs of lines were consecutively numbered.

**Design.** No experimental manipulation was used. The design was correlational in nature and contained three components. The first of these components was to determine whether the free recall subjects organized the pairs. For each subject, both two-way and one-way indices of subjective organization (Bousfield & Bousfield, 1966) were measured, and two-tailed _t_-tests were used to determine whether overall organization scores indicated more observed than expected repetitions of adjacencies. The second component was to determine whether increases in recall over trials were correlated with increases in organization. Mean subjective organization scores and mean recall scores were compared as a function of trials. A Pearson product-moment correlation coefficient was used to make this comparison.

The third component was an examination of the relationship between the data from the triads task and the data from the free recall task. For the free recall task, the proximity matrix contained the number of times, collapsed over subjects and trials, each pair was recalled adjacenty. For the triads task, the proximity matrix contained the number of times, collapsed over subjects, each pair was judged as similar to every other pair. A Pearson product-moment correlation coefficient was used to compare the two proximity matrices.

**Subjects.** The subjects were 20 volunteers from introductory psychology classes. Ten subjects served in the free recall task, and the other ten subjects served in the triads task. All the subjects were tested individually.
Procedure. For the recall task, the subject was told that he would be given a deck of cards and that each card contained a pair of words. He was to go through the deck at his own pace, reading each card aloud, studying it, and then turning it face down. He was not to go back to a card once he had turned it face down. The subject was told that after he turned over the last card, he would be read a three-digit number. He was to count backwards by ones from this number for fifteen seconds. After fifteen seconds, he would be told to stop counting and asked to write down on the first page of his answer sheet, in any order he desired, as many pairs from the list as he could remember. He was to write down both members of a pair on adjacent lines, and he was told to write down both members of a pair before writing down the next pair. After he could think of no more pairs, he was to turn the page of his answer sheet. Then the same procedure would be repeated, as it would for a total of seven trials. The two decks of cards were alternated as stimuli on each trial, and the experimenter shuffled the decks during the interval in which the subject recalled the pairs.

For the triads task, the subject was told that he would be given a deck of cards and that each card contained three pairs of words. His task was to decide, for each card, which two of the pairs were most similar. He was to go through the deck, one card at a time. For each card, he was to read the number in the upper right-hand corner and then read the numbers of the two pairs he thought were most similar. The experimenter recorded the subject’s responses on the answer sheet. The deck was shuffled before being given to the subject.
Results

Organization. The mean number of pairs recalled per trial, collapsed over subjects, was 10.90, with a range from 10.14 to 11.57. For the two-way index of subjective organization, the mean numbers of observed and expected repetitions per trial were 2.03 and 1.55, respectively. Thus, the mean difference between observed and expected repetitions per trial was 0.48, with a range from -0.46 to 1.51. To determine whether these scores indicated significant organization of the pairs, a t-test was used to compare the mean of 0.48 with a test mean of 0. The result indicated that organization approached the accepted level of significance, t (9) = 2.11, p < .10.

For the one-way index of subjective organization, the mean numbers of observed and expected repetitions per trial were 1.13 and 0.77, respectively. Thus, the mean difference between observed and expected repetitions per trial was 0.36, with a range from -0.01 to 0.74. To determine whether these scores indicated significant organization of the pairs, a t-test was used to compare the mean of 0.36 with a test mean of 0. The result indicated that organization was significant beyond the accepted level, t (9) = 3.15, p < .02.

Figure 9 contains the curves for mean recall and mean differences scores for the two-way index of subjective organization as a function of trials. As can be seen from Figure 9, both recall and organization scores showed systematic increases over trials. The correlation between the organization and recall scores as a function of trials was significant, r = .85, p < .05.
FIGURE 9

Recall and Two-Way Subjective Organization Scores

as a Function of Trials in Experiment VII
The conclusion from this section is that subjects can form hierarchical categories and that the formation of these hierarchical categories correlates with recall performance. That the difference scores showed more consistent organization when the one-way, rather than the two-way, index was used, suggests that the recall of categories within a hierarchical node tended to be in a constant order. That is, the one-way index is sensitive to only constant-order adjacencies, while the two-way index is sensitive both to constant-order and changed-order adjacencies. Thus, the present data confirm, to some extent, the observation of constant order recall in Experiment VI.

**Similarities.** The proximity matrices for the recall and the triads data contained 66 cells each, one cell for each possible set of two pairs. As there were ten subjects in the triads task and the number of times each set of two pairs appeared in a triad was ten, the maximum possible cell entry for the triads proximity matrix was 100. As there were ten subjects in the recall task and there were seven trials of free recall, the maximum possible cell entry for the recall matrix was 70.

A Pearson product-moment correlation coefficient was used to compare the proximity matrices for free recall adjacencies and triads judgments. The correlation was statistically significant, $r = .35, p < .01$. Thus, it can be concluded that the formation of hierarchical categories can, at least to a limited extent, be predicted by similarity judgments data.

**Discussion**

As there were four purposes for conducting the experiments reported
in this chapter, four sections will be used to discuss the results. The discussion of the four purposes will be in the same order as they were introduced.

**Similarity.** The data of Kintsch *et al.* (1970) and Experiments III and IV were successfully extended. Kintsch *et al.* found that similarity judgments can predict order of output from lists of categorized words, and in Experiments III and IV similarity judgments predicted order of output from lists of "unrelated" words. In Experiment VI, the correspondence between similarity judgments data and order of output was found when the list was composed of words which might be considered to belong to a single category, and in Experiment VII, the correspondence was found when the list was composed of pairs of words, with both members of a pair belonging to the same category. Thus, the correspondence between order of output in free recall and judged similarity is very general.

In addition, it was found in Experiment VI that, when clustering was measured as conformity to a category structure derived from similarities data, increases in clustering over trials correlated with increases in recall performance. Also, there was some evidence that subjects who were better clusterers were also better recallers. Thus Experiment VI replicated the correlations, which were found in Experiment IV, between recall performance and clustering. However, the stimuli used in Experiment VI were all "mammal" names. Thus, it can be concluded that a model of free recall should not only provide a mechanism for organizing "similar" items, but it should also explain why organizing similar items correlates with recall performance.

**Types of analyses.** In Experiment VI, similarity judgments and order of output were compared using cluster analysis. The data for the
analysis of the similarity judgments were gathered from a number of tasks by Henley (1969) and the measure of relationship in free recall was the relative distance between each pair of items in the subjects' final recall protocols. The correspondence between the cluster structures derived from similarity judgments and from free recall was found to be high.

The cluster analysis programme also allowed an examination of hierarchical clustering in free recall. The hierarchical cluster structure found in Experiment VI appeared to be sensible in that the nodes were easily labelled, but Henley did not report a hierarchical cluster structure which would allow comparison with that derived from the present recall protocols. Also, one possible artifact of the hierarchical cluster structure reported in Experiment VI was that the recall data were collapsed over subjects. To examine this possibility, in Experiment VII categories were defined for subjects in order to determine whether their organization of categories into higher-order categories could be predicted by similarity judgments data. The results of Experiment VII indicated that such predictions could, in fact, be made. The conclusions from this analysis are specific to a model of free recall which invokes similarity as a mechanism underlying organization. The results of the experiments support a model of hierarchically-arranged chunks.

Order of Output

Examination of recall protocols for two individual subjects in Experiment VI indicated that asymptotic recall tended to be in constant order. The constant order of output for the individual subjects in
Experiment VI tended to conform with the similarity judgments data reported by Henley. That is, cluster structures for the two subjects were correlated with the cluster structures reported by Henley. Also, hierarchical cluster structures for the two subjects indicated some degree of sensibility in the arrangement of the clusters.

The observation of constant order of output does have some implications about types of internal structures used to account for order of output in free recall. If a chunking structure is used, some higher-order mechanism for directing order of output of chunks is necessary. Such a higher-order mechanism would be similar to a hierarchical arrangement of chunks. If a multidimensional structure is used, some form of cueing to direct a systematic search is necessary. If a cue is used to describe the location of more than one item, such a multidimensional structure would be similar to a chunking structure.

Type of Structures

The data reported in Experiment VI pointed to the need for further investigation of hierarchically-arranged chunks. Such investigation was carried out in Experiment VII. Experiment VII included three major findings. First, subjects will organize between categories. That is, the evidence indicated that subjects can form hierarchical category structures. Second, the formation of hierarchical categories correlated with recall performance. That is, organization among categories increased over trials as the probability of recall of the categories increased. Third, similarity among categories can be used to predict the organization among categories. That is, similarity judgments data were
significantly correlated with order of output of categories in free recall. Thus, organization of categories into larger units resembles the organization of words into categories.

Although the results of Experiments VI and VII have not eliminated independent categories or multidimensional spaces as possible internal structures to account for order of output in free recall, they do suggest that a hierarchical arrangement is an attractive internal structure. Such a structure, then, will be pursued in the informal model of recall presented in the final chapter.
CHAPTER 6

SUMMARY

In Chapter 2 three questions which have formed the basis for the present investigation were asked. These were: (1) Given that organization affects recall, at what stage of the recall process is the effect felt? (2) Do the same processes underlie both clustering and subjective organization? and (3) Can a description of order of output be made which would reflect on the internal structure used to generate such output? In this chapter the results of the experiments will be reviewed in terms of the three questions. The question concerning clustering and subjective organization will be discussed first because most of the present experiments are relevant to this question.

Clustering and Subjective Organization

The present dissertation is unusual in that the research has not appeared to concentrate on a specific aspect of free recall, nor has it been designed to test a specific contention or hypothesis. Yet, the research can be regarded as having made a specific point or as having an underlying theme. That is, various types of organization seem to represent the same processes. In terms of the present experiments, this point may be regarded as having three subpoints. First, various manipulations of organization have asymmetric affects on recognition and recall. Second, there are correlations between similarity judgments data and order of output in free recall which extend over categorized and unrelated lists. Third, the amount of organization, as measured by a variety of techniques, can be used to predict recall performance. These three
subpoints will be discussed separately.

Organization and Recognition. Data reported by both Kintsch (1968) and Bruce and Fagan (1970) have shown that varying the strengths of category associates, which affects recall performance, does not affect recognition performance. Thus, if the manipulation used by these experimenters is regarded as a manipulation of organization, it can be argued that the organization of a categorized list may have asymmetric effects on recall and recognition.

Experiments I and II were designed to extend Kintsch's and Bruce and Fagan's findings to lists which did not have an a priori structure. In Experiment I, instructions to use visual imagery to organize items in their input order were found to lead to a significant increase in recall performance but not a significant increase in recognition performance. In Experiment II, instructions to rehearse overtly the item being presented were found to decrease recall performance but to have no effect on recognition performance. Thus, the findings of Kintsch and Bruce and Fagan have been replicated with different types of materials and different manipulations of organization.

Organization and Similarities. The data gathered in Experiments III, IV, VI, and VII are all consistent with the data gathered by Kintsch, Miller, and Hogan (1970). Kintsch et al. observed correlations between judged similarity and order of output in free recall when the materials had an a priori structure. Correlations between judged similarity and order of output were also observed in Experiment III, IV, VI, and VII, but other types of materials were used.
In Experiments III and IV, the stimuli were "unrelated" words. In Experiment III, there was a nine-item list, and number of adjacencies of pairs was the measure of order of output; similarity judgments were from a triads task. The measure of relationship was a correlation between the proximity matrices. In Experiment IV, there was a 40-item list and number of adjacencies was again the measure of order of output; similarity judgments were from a card-sorting task. Correspondence between cluster structures derived from the proximity matrices was used to measure the relationship.

In Experiment VI, the stimuli were 30 items which might be considered to belong to a single category. Relative distance between each pair of items on the subjects' final recall protocols was the measure of order of output. Similarity judgments data had been gathered in a previously-published study (Henley, 1969). Again, correspondence between cluster structures derived from the proximity matrices was used to measure the relationship.

In Experiment VII, the stimuli were 12 pairs of words, with both members of a pair belonging to the same category. Number of adjacencies between the pairs was the measure of order of output, and a triads task was used to measure similarity. The measure of relationship was a correlation between the proximity matrices.

Experiments III, IV, VI, and VII all showed that similarity judgments data can predict order of output in multtrial free recall. In Experiment V, similarity judgments data were used to predict order of output in single-trial free recall. The stimuli were the 40 unrelated
words used in Experiment IV, and the category structure was derived from the sorting data in that experiment. In Experiment V significant clustering according to the category structure was observed in single-trial free recall. As one of the operational distinctions between subjective organization of unrelated words and clustering of categorized words is that the former cannot be measured in single-trial free recall, the results of Experiment V serve to reduce the distinction between the two types of organization.

A conclusion that judged similarity can be used to predict order of output appears to be well supported. Correspondences have been found over different types of materials (categorized lists and unrelated lists), a large range in number of items (9-40), different types of judged similarity (card-sorting and triads judgments), two measures of order of output (adjacencies and relative distance), and two types of comparisons (correlations between proximity matrices and between cluster structures). It seems appropriate to conclude that the relationship between judged similarity and order of output in free recall is robust.

Organization and Performance. The major findings which support the contention that increases in organization are necessary for increases in recall performance are the correlations which have been observed between measures of organization and amount recalled (e.g., Tulving, 1962; Mandler, 1967). Regardless of whether organization is necessary for recall, the present experiments have yielded further evidence that organization correlates with recall.

The usual measures of organization are clustering scores, which are
used when the list is composed of categorized words, and subjective organization scores, which are used when the list is composed of "unrelated" words (see Chapter 1). In Experiment IV, the list was composed of unrelated words, but clustering scores were used to measure organization. That is, the results of the cluster analysis of the similarity judgments data were used to treat the list as if it were composed of categories, and correlations between recall performance and clustering scores were determined. Correlations between increases in recall and increases in organization as a function of trials and between total recall and total organization scores over subjects were significant.

In Experiment III, the list consisted of associates to a single category name. Again, results of similarity judgments data, collected by Henley (1969), were used to treat the list as if it were composed of associates to multiple categories. Clustering was scored according to the amount of conformity of the subjects' recall protocols to these categories. A high correlation between increases in clustering scores and increases in recall performance over trials and a positive, although not statistically significant, correlation between total recall and total organization scores over subjects were observed.

In Experiment VII, the list was composed of unrelated items and, as is typical with unrelated items, subjective organization scores were used to measure organization. However, the unrelated items used were unusual in that each item was a pair of words, with both members of a pair belonging to the same category. Even with these pairs, increases in recall over trials were found to correlate with increases in organization.
Another finding related to organization and performance was the simulation of the Shapiro and Bell (1970) study in Experiment IV. Here, relationships between organization and recall over trials for high, middle, and low organizers were observed. In accordance with the Shapiro and Bell results, recall increased over trials for all three groups, and high organizers, but not low organizers, showed corresponding increases in organization scores. However, although in both Experiment IV and the Shapiro and Bell study unrelated words were used, organization was measured as conformity to a category structure derived from similarity judgments data in Experiment IV, while organization was measured by subjective organization scores in the Shapiro and Bell experiment.

Thus, the results of experiments which used unusual types of materials and/or unusual combinations of organization measures and types of materials showed positive correlations between organization and recall performance. It can be concluded that, over a range of materials and organization measures, increases in organization over trials predict increases in recall performance. It can also be concluded that differences in organization among subjects can, to some extent, predict differences in recall performance.

Conclusion. In Chapter 1, the clustering of categorized words was compared with the subjective organization of unrelated words. Differences between these two types of organization were found to be limited to the type of material used and the unit of measurement employed. Four manipulations were discussed which produced differences in organization
scores and recall performance independently of whether categorized or unrelated lists were used. One possible conclusion is that clustering and subjective organization represent the same processes.

In this section, three more commonalities between the clustering of categorized words and the subjective organization of unrelated words were noted. First, manipulations of organization of unrelated words and categorized words may produce asymmetric effects on recognition and recall. Second, similarity judgments can predict order of output in the free recall of both unrelated words and categorized words. Third, clustering measures can be used to predict recall performance when the list is composed of unrelated words. Thus, the data reported in this dissertation can be viewed as supporting the notion of common processes underlying two operationally distinct types of organization.

Storage and Retrieval

The observation in Chapter 3 of asymmetric effects of organizational manipulations on recall and recognition performance was interpreted as evidence that organization affects the amount retrieved but not the amount stored. The assumption underlying this interpretation was that recognition measured storage and that recall measured storage plus retrieval. Again, this is not the appropriate place for rehashing the argument of whether retrieval is involved in recognition, and the reader is referred to papers by Kintsch (1970) and McCormack (1972) for data pertinent to this argument.

It should be pointed out that one manipulation of organization has been shown to affect recognition performance. That manipulation
is one of presentation order, such that blocked presentation of
category members results in better recognition performance than random
presentation of category members (D'Agostino, 1969; Jacoby, 1972). A
possible explanation for the discrepancy between this manipulation and
the manipulations used in Chapter 3 and by Kintsch (1968) and Bruce
and Fagan (1970) will be advanced in the next chapter.

Besides reflecting on the storage-retrieval question, the results
of the experiments reported in Chapter 3, particularly Experiment I,
have another implication about the nature of organizational processes.
That implication is that organization can be a control process. Ex-
periment I showed that subjects will use instructions to form visual
images in order to organize a list of unrelated words, and that using
these instructions will lead not only to increases in recall performance,
but also to increases in clustering according to input position. As
the list used in Experiment I did not have an a priori structure and
presentation order was randomized, it is likely that the organization
resulting from use of the imagery instruction was not the organization
that the subjects would normally use. Thus, it can be concluded that
there is some flexibility, which is under the subject's control, in
organizational processes.

Order of Output and Internal Structure

The major finding of the experiments reported in Chapter 5 in-
volved hierarchical organization. In Experiment VI, the cluster
analysis of the group recall data produced a hierarchical cluster
structure which was intuitively pleasing. That is, the combination
of categories into higher-order categories appeared reasonable, and the nodes of the higher-order categories could be easily labelled.

In Experiment VII, hierarchical clustering was examined more directly than in Experiment VI. The stimuli were two-word categories, and organization was measured as the combination of these two-word categories into higher-order categories. The data indicated that the subjects were organizing into higher-order categories and that this organization increased over trials.

A third finding involving hierarchical clustering was the analysis of order of output for two individual subjects in Experiment VI. Both of these subjects recalled in a constant order over the last four trials. In the context of a clustering model, constant order of output would necessitate a mechanism for ordering the recall of various categories. Such a mechanism could be similar to higher-order nodes in a hierarchical clustering model.

Thus, the experiments reported in Chapter 5 indicate that hierarchical clusters may be an adequate representation of asymptotic free recall protocols. This does not show that a hierarchical structure is the appropriate structure to represent the set of relations among words which results from a single subject learning a free recall list. However, it appears convenient to explore the possibility that there is a relatively direct correspondence between the structure of organization observed in sets of free recall protocols and a single subject's internal representation which produces that structure. Thus, a model employing a hierarchical internal structure will be pursued in the next chapter.
CHAPTER 7
THEORETICAL NOTES

This chapter contains two major sections. In the first section an informal model of free recall is proposed. In the second section, the data reviewed in the first chapter is described in terms of the model and some experiments suggested by the model are briefly described.

Towards a Model

In this section, some thoughts about a model of free recall are presented. The model is described in three subsections. In the first subsection, an internal structure of organization which is consistent with the description of output order extracted from the results of the present experiments is described. In the second subsection, the process by which the internal structure develops is described; the process is similar to that outlined by Kintsch (1970). In the third subsection, the relationship between organization and tests of recall and recognition is described, and a possible explanation of why organization may have asymmetric effects on recognition and recall is advanced.

Internal Structure

In relation to free recall, the term structure may be thought of as having three referents. The first of these is to list structure, that is, the structure that may be manipulated by the experimenter. For instance, one type of list structure may be described as items belonging to a number of mutually exclusive categories. Another type of list structure may be described by pairwise measures of judged similarity among the items. The second type of structure is that found in subjects'
recall protocols, and in the present paper this type of structure has often been referred to as organization. For instance, recall structure may be described as the amount of conformity in the subjects' recall protocols to list structure. Another example of recall structure may be described as a given subject's tendency to repeat specific adjacencies from one trial to the next.

A major thesis of the present paper has been that, to some extent, recall structure reflects list structure (see Chapter 4). That is, the subject processes items from the list, and the product of this processing is a mechanism which generates recall structure. In this section, the mechanism which results from the processing is of interest, and this mechanism is the third type of structure, internal structure. As a mechanism, internal structure should be thought of as a structure which results from the subject's processing of the items on the list and from which the recall structure is generated. As a structure, internal structure should be thought of as the subject's representation of the relationships, or patterns of sets of relationships, among the items on the list to be recalled. In the present section, these relationships are described.

A model of internal structure should account for the data concerning order of output. That is, the model should explain the observed relationship between order of output and judged similarity. Also, the model should produce an order of output which can be described as a hierarchical combination of categories. Finally, a model of internal structure should allow the development of a highly
consistent order of output of categories and of items within categories.

The present model of internal structure is a hierarchical combination of categories and items within categories. The hierarchical structure develops because of some limitations on processing which are discussed in the next section. Constant order of output of categories and of items within categories is related to a concept of associative strength between higher-order category codes and category codes and between category codes and items. Strength of association, here, may be thought of as the ability of a higher-order category code to elicit a category code and as the ability of a category code to elicit the items within that code. Because a code is described as a characteristic common to the items within that code, a code serves as a mechanism by which order of output is related to judged similarity.

The present model of internal structure is one of a static structure. A static structure may be thought of as the end product of the subject's processing of the items on the list, and it is the structure from which asymptotic recall is generated. Within this structure there is some limit to the number of categories and to the number of items within a category. If the number of categories exceeds this limit, it is likely that the subject combines categories into higher-order categories. This combination keeps the items within a subcategory distinct so that all of the items within one subcategory are recalled before all of the items within a second subcategory. That such a structure is possible was suggested by the results of Experiment VII. The process by which this structure develops and the
limitations which result in the formation of "internal" categories are described in the next subsection.

A category is considered to be some series of associations rather than as a "chunk" in Miller's (1956) sense. A "chunk" implies some process of recoding or relabelling two or more items into a larger item. This "unitization" of items leads to increased recall capacity; however, along with this increase, there is a loss of information about the original stimuli. This loss of information with unitization was clearly demonstrated by Horowitz and Manelis (in press). Their subjects were presented with three types of adjective-noun pairs: unitized phrases (such as hot and dog which could be recoded as hotdog), meaningful phrases (e.g., blue and shoes), and anomalous phrases (e.g., hot and shoes). They found that recall was best for the unitized pairs, but that recognition for individual items was worst for these pairs. Their conclusion was that although unitization resulted in greater recall, it caused a loss of information about the individual items.

That typical organization in free recall does not involve a loss of information about individual items is fairly clear. Recognition memory for individual items has not been shown to be worse, and is sometimes shown to be better, following manipulations which are hypothesized to increase organization. The discussion of previously published experiments and some original experiments in Chapter 2 includes a variety of manipulations of organization, none of which depressed recognition performance. Thus, there is some empirical evidence to reject the notion
of chunking as descriptive of categories which are formed in free recall.

Other reasons to reject the chunking notion involve the problems of analogy from the immediate memory paradigm to the free recall task. The reader is referred to Chapter 2 for discussion of these problems.

Although chunking does not seem to be an accurate description of internal structure, the idea that a subject produces a code is attractive. The cluster analysis of the data presented in Experiment VI indicated that the nodes were easily labelled, and informal reports of subjects who participated in that experiment indicated that they were actively naming their categories. Thus, although the evidence does not necessitate a coding process, it is not unreasonable to postulate such a process. In this way, categories may be described by codes, and higher-order categories may be described by higher-order codes. The structure, then, may be referred as a coding hierarchy.

One possible description of a coding hierarchy is multiple associations between a category name and the members of that category. Recall of items within a category is ordered by some probabilistic function of the varying strengths of association between the category code and the items within that category. For instance, one possible decision rule is to have the subject, upon recall of a category code, recall the strongest associates to that code before recalling the weaker associates. Thus, one would expect certain orders of recall of items within categories to be more prevalent than other orders. As
an illustration, such prevalence was investigated for the three jungle cats used as stimuli in Experiment VI. Here, the author's intuition was that free association to the category name jungle cat would have lion as a higher associate than tiger which would be a higher associate than leopard. In fact, when these three items were recalled adjacently in Experiment VI, the order lion, tiger, leopard appeared 41% of the time, while the other five orders appeared a total of only 59% of the time ($X^2(1) = 32.9$, $p < .001$). A similar process is postulated to describe the combination of codes into higher-order codes. Further, the highest-order codes are associated with a code such as list-membership, and the highest-order codes are recalled in an order depending upon the strength of association with list membership. Strength of association is defined in terms of two constructs. First, associative strength between an item and code can exist for the subject prior to the experiment. Second, the strength may vary as a function of the amount of practice during learning of the list to be recalled.

There are other possible representations of a structure similar to that described above. For example, categories could be cued by a category exemplar, rather than by a category code, and that exemplar would exist in some series of item-to-item associations with other

---

1 The author's intuition does not necessarily reflect normative associative strength, and no norms exist to the category jungle cat. However, lion is a stronger associate to wild animal than is tiger which is a stronger associate than is leopard (Shapiro & Palermo, 1970).
members of that category. The types of associations could be described so they could account for constancy in order of output. It would be difficult to discriminate between this type of representation and a representation of category codes. As Tulving (1968) has pointed out, items which share a common code probably are associates of each other, and items which are associates of each other probably share a common code.

**Processing**

A model of processing should account for the development of the internal structure described in the previous section. There should be a reason why the subject forms categories, and this reason, in the present model is related to some limitations on processing. The development should involve similarity among items, so that more "similar" items are likely to be contained in the same category (see Chapter 4). The model should also account for the serial position curve found in free recall (see Chapter 1), and it should describe different processes for recognition and recall (see Chapter 3).

The description of processing involves two memory stores, both of which are related to the internal structure described in the previous section. The first of these stores is a long-term memory store. Here, the subject has pre-experimental characteristics of items at his disposal. The second store is a short-term memory store and is called a limited capacity processing mechanism (LCPM). This store not only accounts for the short-term memory data described in Chapter 1 but also is the center for processing of the items. The stores are
related to the internal structure in that information from long-term memory is processed in the LCPM and this processing produces the internal structure. The internal structure is labelled the free recall memory structure or FRMS, and it was described in the previous section. Descriptions of long-term memory and the LCPM follow.

The LCPM is similar to the "buffer" used by both Atkinson and Shiffrin (1965, 1968) and Kintsch (1970). However, it is afforded properties that these theorists did not use, so the term buffer is not employed. The LCPM is viewed as a short-term memory input-output mechanism which has four major properties. First, in input, recent items may enter and reside in the LCPM. There is a limit to the number of items which can reside in the LCPM, and the variables which affect the LCPM in this way have been described by Atkinson and Shiffrin (1965, 1968). This first property accounts for the short-term memory component in free recall. The second property of the LCPM is that "cognitive work" is performed in it. This cognitive work is similar to that described by Kintsch (1970). The assumption here, is that when an item is in the LCPM, its long-term memory characteristics are transferred to the LCPM and are available for processing. The processing will be described later in this section, and the purpose of the processing is to find a code which is shared by two or more items in the LCPM. The third property of the LCPM is the strengthening of associations between a code and items within the code so that a code is more likely to elicit its associated items. Here, after a code has been found for two or more items, the more time those items
reside in the LCPM, the stronger the strength of association between code and items. Associations may be strengthened by some process such as contiguous rehearsal (see Rundus, 1970), although it is not necessary to describe the exact nature of the strengthening process at this time. The fourth property of the LCPM involves recall of items represented in the FRMS. Here, items represented in the FRMS are transferred to the LCPM for output. The limited capacity of the LCPM requires that only a certain number of items may be transferred at any one time, and, thus, a limited number of categories and items per category may be derived. This fourth property will be pursued in the discussion of recall in the next section.

In agreement with the theories of Kintsch (1970) and Anderson and Bower (1972), the representation of an item for processing consists of a list of traits, characteristics, or markers rather than the item itself. The markers represent a variety of types of characteristics of the items, such as superordinates, item-to-item associations, and acoustic and formal properties. That properties other than common superordinates will serve as a basis for organization has been demonstrated in a number of experiments, some of which were discussed in Chapter 1.

Although an item may have many markers, not all of these markers are equally available. This notion of differential availability serves as the basis for the processing model. Availability, here, may be thought of in associationistic terms such that a more available marker is more easily elicited by the item. Which types of markers are more
available varies as a function of the specific item and the item's context. In the first case, the differential availability of an item's markers may result from the subject's pre-experimental experience with that item. For instance, the marker animal may be rapidly found upon presentation of the word dog, where formal properties, such as the initial sound is d, will be less rapidly found. Conversely, the marker the initial sound is z may be more rapidly found for an item such as xylem, an item with which the subject may not have had much "semantic" experience. Thus, it could be predicted that less frequent words would more likely cluster according to formal or acoustic properties than more frequent words.

In the second case, availability of markers varies with the item's experimental context. That is, certain experimental manipulations may make a marker more available. For instance, instructions about how to organize affect the availability of markers. When told to organize alphabetically, a subject will use the markers associated with the items' initial letters rather than the markers associated with the items' meanings. A second example of contextual availability takes the form of a prediction. For instance, if early in the list, the subject is presented oboe and folk adjacently, it is likely that he would find markers which correspond to music to use as a code. (Marker will be used to define information about an item which exists in long-term memory; after a common marker is found, it will be referred to as a code. A code, here, is what was described in the section on internal structure.) If later in the list he is presented singer, the likelihood
that he would find the marker **music** rather than **occupation** should be increased. Processing which can account for this type of contextual availability is described later in this section, and a direct experimental test of the prediction is suggested in the last section.

The discussion of the cognitive work which is performed in the LCPM only concerns itself with the case where the LCPM is filled. The LCPM is assumed to be filled except at the beginning portions of the serial order of presentation, so the description should hold for most of the list presentation. Descriptions of processing for the first few items, particularly as the processing relates to the primacy effect in single-trial free recall, can be found in Rundus (1970).

When an item which is presented first enters the LCPM, its list of markers is extracted from the subject's long-term memory. The list of markers is found by some process of matching, and this matching involves direct access to the item. That is, the subject need not engage in a complex search to find that item's markers (see Kintsch, 1970). Two operations are accorded this list of markers. First, the list is given a time tag similar to that described by Kintsch (1970). The time tag acts like a marker in that it is elicited by the item, and it is available for tests of recognition. Second, the tagged list of markers is processed in the LCPM. In the LCPM, this list of markers is compared with other lists of markers already in the LCPM. The comparison involves a "reading off" of the markers for the new list and the markers are read in some order, with the most available markers being read first. Each marker that is read from the new list of markers
is compared with the other marker lists in the LCPM. This comparison with other items may proceed in some serial manner, with the initial comparison being made with the most recent item to enter the buffer. If the most available marker from the new list fails to be found in the other lists of markers, the next most available marker from the new list is extracted, and the process recycles.

If a match of markers is found, the marker which is shared in common with the two items is compared with the remaining marker lists in the LCPM to determine whether that marker is also shared by any of these remaining items. Then, the associations between the common marker and the items are strengthened; the contextual strength of association between the items and the common marker is a function of the amount of time, after the common marker is found, that the items reside in the LCPM.

When a marker common to two or more items is found, that marker becomes a code for those items, and, thus, the code and the items are said to be represented in the FRMS. The strength of association between the code and its items is a function of both the availability of the code and the amount of time the items and the code have resided in the LCPM.

In some cases an item will enter the LCPM and no marker common to another item in the LCPM will be found. That is, the more available markers for that item will not be common to other items in the LCPM, and the item may not reside in the LCPM long enough for some less available markers to be processed. In this case, when the item leaves
leaves the LCPM, its time-tagged marker list is available to the subject for tests of recognition.

After the first common marker has been found and along with its associated items is represented in the FRMS, the processing of new items entering the LCPM proceeds in a slightly different manner. First, as was the case when no codes have been established, the marker list for the item is extracted from the subject's long-term memory, is time-tagged, and enters the LCPM. Then, however, this list of markers is not compared with the marker lists for other items in the LCPM; rather, it is compared with the codes for items which are represented in the FRMS. If the marker list for the new item is found to contain one of these codes, the items in the FRMS which share this code are brought back into the LCPM. Here, the associations between items and codes are strengthened.

If the new item does not contain a marker that is common to a code for items represented in the FRMS, the comparison process proceeds as if there were no items represented in the FRMS. That is, the marker list for the new item is compared with the marker lists for other items in the LCPM. Regardless of whether a comparison leads to a match in either the FRMS or the LCPM, when the item leaves the LCPM, its time-tagged marker list is available for tests of recognition.

The strength of the category code is defined as the strength of association with a higher-order code, that code being list-membership. The strength of the code is determined by the amount of time that code resides in the LCPM. Thus, the strength of a category code should
be some monotonically increasing function of the strengths of the items associated with that code.

The hierarchical nature of organization may result in one of two possible ways, and it is not unlikely that both ways occur in typical free recall situations. First, in initial stages of free recall learning, some code which describes a number of items in the list may enter the FRMS. As the list is presented, many items may become associated with that code, and the total number of items so associated may exceed the capacity of the LCPM. That is, during recall, the items within a code are assumed to channel through the LCPM, and there is a limit to the number of items within a code which can be transferred to the LCPM (see the description of recall in the next section). Thus, the subject may limit the number of items within a code by finding new codes which describe only a subset of the items. To do this, as free recall learning progresses, he will compare marker lists of items which share this common code in search of a more available code. This search proceeds in the same manner as that which led to the initial code, and the strengthening of associations with the new code will also proceed in the same manner. In this way, an item will be associated with more than one code, and the codes will be associated with that item in a hierarchical manner. Thus, this first way in which a hierarchy develops can be described as development from higher levels to lower levels.

It is also possible that hierarchies may develop from lower levels to higher levels. Within the structure of the proposed model, it is possible that the subject will form more than the number of codes that
can be processed in the LCPM. That is, codes are assumed to channel through the LCPM and there is a limit to the number of codes which can be transferred to the LCPM. In this case, on subsequent trials, the subject could process the codes as if the codes were items. That is, comparison of codes could lead to the matching of a marker shared in common by two or more codes, and these codes could be combined and associated in the same way items within a code become associated. Thus, as an item can become associated with a code and that code can become associated with a higher-order code, a hierarchical structure can develop. That a hierarchy can develop from lower levels to higher levels was demonstrated in Experiment VII.

Recognition and Recall

A model of output should describe why manipulations of organization may have asymmetric effects on recognition and recall (see Chapter 2), but it should also explain when manipulations (i.e., blocking) may have symmetric effects. A model of output should also describe the transfer of information from the internal structure to the recall protocol.

In the present section three types of recognition memory failures are described. One of these, homographic encoding, can be affected by manipulations of organization. The other two are failure to store and decay or interference. Recall is described as transfer from the FRMS through the LCPM. The description of recall accounts for consistent order of output. The description of recognition memory proceeds first.
The description of recognition memory is similar, in some aspects, to that presented by Winograd and Raines (1972). Two basic assumptions are necessary for this description. First, some items are assumed to have two or more lists of markers in memory. This assumption may best be explained in terms of homographs, i.e., words with two or more distinct meanings. For instance, orange may have a list of markers which corresponds to orange as a color, and it may have a second list of markers which corresponds to orange as a fruit. Although the marker list for orange as a color may contain a reference to the marker list for orange as a fruit, the two marker lists are distinct. It should also be assumed that for an item with two or more marker lists, one of these marker lists is more available. That is, when the item is presented, there are differential probabilities of extracting the various marker lists from memory. Operationally, the more available marker list for known homographs can be defined in a free association paradigm. For instance, if orange is a stimulus for free association, and the subject responds apple, it is likely that orange as a fruit is more available than orange as a color. If he responds red, it is likely that orange as a color is more available. The availability here is assumed to be probabilistic; that is, if the subject is presented the same homograph on two different trials, there is a chance that he may extract different marker lists on each of these trials. It should be pointed out that some recent studies have collected norms of association to homographs (e.g., Kausler & Kollasch, 1970).
The second assumption is that recognition memory does not involve complex search processes. An item which is presented in a recognition memory task is directly found in the subject's long-term memory. Here, the subject checks the marker list of that item for its time-tag. If the time-tag indicates that the item was on the list, he can correctly recognize it. If no time-tag is found, the subject reports that the item is a new item. Thus, some of the failures to recognize items which were on the list can be attributed to failures to time-tag that item in the LCPM.

Two other types of recognition memory failures can be identified. First, the time-tags are given some decay (or interference) characteristic. This decay characteristic is essential given that recognition performance does suffer over delays from presentation to test (see e.g., Wickelgren & Norman, 1966). Also, a decay is necessary in order to provide for false alarms, i.e., responses of "old" to distractor items. That is, if the subject is responding to decaying time-tags at a low enough criterion, he may respond to items which have not been presented. A more complete discussion of time-tag decay and recognition is presented in Kintsch (1970).

Second, recognition memory failures occur as a function of the probabilistic encoding of homographic words. For example, on the study trial, the subject may see orange and tag the list of markers which corresponds to orange as a fruit. On the test trial, with some probability he may see orange and extract the list of markers which corresponds to orange as a color. This list of markers will not contain a time-tag
indicating that orange was on the list, and thus, the subject will fail to correctly recognize orange. That such failures to recognize as a function of changed meaning occurs has been demonstrated in a number of studies which have manipulated the meanings of homographs on study and/or test trials (e.g., Light & Carter-Sobell, 1970).

The presentation of items in blocked fashion should bias which marker list is chosen. Also, the use of category distractors should bias which marker list is chosen for recognition. For blocked presentation, both presentation and recognition bias should be the same. For spaced presentation, there should be no bias at presentation, although there will be a bias at recognition. Thus, blocked presentation should produce better recognition performance.

Manipulations of organization other than blocking have not been found to affect recognition performance, probably because they have not differentially biased meaning between groups with different levels of organization. For example, in Experiment I, the items were "unrelated" and the distractor items were randomly chosen. In Experiment II, although the distractor items belonged to the same categories as the presented item, there were no systematic relationships between the items in the presentation list. Thus, both rehearsal groups should experience the same difficulty in terms of the distractor items. Choosing distractors from the same categories as the presentation items should lead to increased difficulty in recognition (cf. Bower, Clark, Lesgold, & Winzenz, 1969).

While the recognition process involves direct matching with time-
tagged marker lists, the recall process involves retrieval from the FRMS which has been hypothesized to be a hierarchical structure of the presentation items. The hierarchy is assumed to be a set of associations such that list-membership is associated with category codes and category codes are associated with either lower order codes or with the items. Recall from this structure is assumed to follow a path from the top to the bottom, and probability of recall of both codes and items within codes is assumed to be a function of the associative strength.

Recall of the items for this structure is assumed to channel through the LCPM. First, especially early in learning, the LCPM is emptied of recently presented items (see e.g., Kintsch, 1970). Then, the code list-membership is transferred to the LCPM along with the highest-order codes. The number of codes which can be transferred at any one time is limited by the capacity of the LCPM. Then one code, probably the highest associate of list-membership, is chosen. Remember that the strength of association of list-membership to a code is a function of the amount of LCPM time that the code has had along with its associated items. When this code is chosen, the items within that code (or lower order sub-codes) are transferred to the LCPM. These items are recalled in some order, possibly dependent upon the strengths of association of the code to the different items. Here, strength of association is a function of both amount of LCPM time and the availability of the code. Availability of a code has been described earlier.

After the LCPM has been emptied of items within the strongest
code, the next strongest code is used to transfer items from the hierarchy to the LCP! and the process is assumed to recycle throughout the hierarchy. In this way, recall may approach a constant order. As the recall of the code is assumed to precede recall of items within that code, failure to recall a code will preclude recall of its associated items.\(^2\)

**Experiments and Predictions**

In this section, the model is discussed in two ways. First, an attempt is made to show how the model can handle the free recall data discussed in the first chapter. Second, some possible experimental tests of the model's assumptions and of predictions that can be made from the model are suggested.

**Evaluation**

The model makes no distinction between the clustering of lists with an \textit{a priori} structure and the subjective organization of "unrelated" words in terms of underlying processes. The major theme of the experiments reported in this dissertation has been that, regardless of the types of items used, similar manipulations of organization produce similar results in performance.

The model can handle the data which indicate that lists with high associative structure are recalled, but not recognized, better

---

\(^2\) It is possible, as Mandler (1967) suggests, that an item could have more than one representation on the hierarchy. This possibility could be handled by the model; however, it will not be pursued.
than lists with low associative structure (e.g., Kintsch, 1968). The discussion of the recognition process indicated that, unless the manipulations of organization produced differential homographic encoding for different groups of subjects, there should be no change in recognition performance. If both high and low associatives lists are presented in a random order, it is difficult to see how encoding could be affected; again, if both are presented in blocked order, encoding bias should be approximately the same for both groups.

The notion of marker availability, however, can explain why highly structured lists are better recalled. Highly structured lists are assumed to contain more available common markers than lists with little structure. The advantage to the subject, then, can occur at two stages of the recall process. First, as more available common markers take less LCPM time to find, the subject will have more LCPM time to strengthen the associations between the items and the codes. Second, as the strength of association between a code and its associated items has some pre-experimental value, more available markers should be better at eliciting their associated items.

As the model does not distinguish between the clustering of words with an \textit{a priori} structure and the subjective organization of "unrelated" words, the remainder of this section will treat the two types of organization as the same. The description that follows is ordered in the same way as the review of the data in Chapter 1.

\textbf{Order of Input.} If items with more available common markers reside in the LCPM at the same time, there will be less time necessary to
find common markers and more time available for strengthening of associations. As blocked presentation should increase the probability that items related by more available common markers reside in the LCPM at the same time, it should be predicted that blocking would lead to greater associative strength between codes and items within codes. Thus, recall performance should be improved. Higher organization scores can also be predicted. In random presentation, subjects may extract codes which are less available than those built into the list structure. Further processing of items within these codes, then, may be necessary to limit the size of categories (see the discussion of how organization develops from the bottom of the hierarchy to the top). With this further processing, the subject's organization will change. In blocked presentation conditions, no such further processing may be necessary. If it is, it is likely that the processing would proceed from the bottom of the hierarchy to the top, and little change in adjacencies at recall would occur.

Presentation Rate. The arguments concerning presentation rate follow those concerning blocked presentation. That is, slowing presentation rate should increase the LCPM time available to the subject and thus, more time would be allowed for strengthening of associations. In this way, increased recall and increased organization scores can be predicted.

Cueing. The model allows some independent "forgetting" of the code and its associated items. When manipulations insure that the strengths of association between codes and their associated items are about the same over all categories, such independent "forgetting" would be
expected. Presentation of code names along with items or blocked presentation of items within a code should, to some extent, equalize strengths of association between codes and items within codes. Thus, if a code is "forgotten" and such presentation conditions are used, when the subject later "remembers" that code, the number of items that he recalls should not be less than the number that he would have recalled had he not forgotten the code.

Codes may be "forgotten" if the number of codes exceeds the number which can be transferred from the FRMS to the LCPM. The purpose of a cue, then, is to bypass the necessity to transfer more than one code from the FRMS to the LCPM at any one time. A cue is envisioned as providing direct access to the FRMS, much in the same way that after the FRMS develops, the subject compares markers lists for new items with the category codes in the FRMS.

As the cue is directly compared with the category codes, it is necessary that, in order to provide a "match," the cue be identical to the category code. Thus, for cueing to be effective, manipulations which insure that the subject uses codes which will eventually be used as cues are necessary. Blocking, presenting the cue at presentation, and asking the subject to name his categories are such manipulations. That these manipulations are effective in producing additional recall with cueing was discussed in the first chapter.

Number of Categories. As the model postulates that there is a limit to the number of codes that can be transferred from the FRMS at any one time, number of categories should be an effective variable.
However, the optimal number of categories should also be a function of list length, as the number of items within a category is also assumed to have some limitation. Thus, it is not surprising that an effect of number of categories has been found but that a function describing this effect is not clear. At present, the author will not attempt to formulate such a function.

**Inhibition.** The model which has been presented has not been constructed to handle learning of two or more lists. However, it is not unlikely that it could be modified for such situations. Elsewhere, it has been suggested that list differentiation may account for some effects of proactive inhibition (Schwartz and Humphreys, 1972b), and this suggestion will be pursued in refinements of the model.

**Individual Differences.** If, as Earhard (1967) hypothesizes, individual differences in free recall represent individual differences in the ability to form associations, the present model can handle such differences. That is, recall is viewed as sets of associative bonds between items, category codes, and higher-order codes, such that individual differences in ability to form these bonds will result in differences in both recall and organization scores. However, this explanation really avoids the more interesting question: why are there individual differences in ability to form associations?

**Short-term Memory.** As the present model employs a buffer-like construct (LCPM), such findings as the large recency effect are easily explained. The model would handle short-term memory phenomena in the same way as Kintsch's (1970) model.
Future Research

The model has many aspects which are speculative. Many of these might be open to experimental investigation. Also, the application of the model to the data may suggest new lines of investigation. In this section, some experiments suggested by the model will be discussed. There is no contention that all, or even a good percentage of, possible experiments may be represented. Rather, a number of studies which interest the author will be briefly described.

Structure of Organization. The data indicated that a hierarchical category arrangement is a good description of the structure of organization. However, there are no data which prove this contention. So far as the author knows, there is no way to prove or disprove any representation of structure for free recall memory. However, an interesting question can still be asked: is a hierarchical arrangement the most efficient structure for recall? A possible approach to answering this question would be to have different groups of subjects learn concept identification tasks with solutions being based on clustering, hierarchical categorization, or multidimensional designs. The items would be words and they would be the same for all groups, and the groups would be tested on recall of the words. If a suitable design can be found, the author would like to pursue this line of research.

The notion of constant order recall of categories and items within categories as a function of strength of association can be tested. As strength of association is defined in two ways, two manipulations
suggest themselves. First, pre-experimental strength of association of items to a code can be manipulated, within a category, by using low and high category associates. It would be predicted that when the category is recalled, high associates would tend to be recalled before low associates. Second, contextual strength of association may be varied by the amount of LCPM time afforded an item and its category name. Here, an item which belongs to a category might be repeated twice in the same list presentation in order to give it more LCPM time. It would be expected that, when the category is recalled, the item with multiple presentations will tend to be recalled first, provided that it is represented at only one location in the structure.

**Processing.** Three lines of research are suggested by the discussion of processing. First, the notion that processing proceeds in the LCPM in a serial manner makes a prediction about presentation order. Here, the amount recalled should decrease when related items reside in the LCPM at the same time, but have not been presented adjacently. In order to test this prediction, a list of two-word categories could be used. Items within the same category would be presented either adjacently, with one intervening item, or with two intervening items. It would be predicted that the number of categories recalled would vary inversely with the number of intervening items. If the prediction is not substantiated, the model may have to be modified to include parallel processing in the buffer.

The second line of research involves whether all marker comparison takes place in the LCPM or whether there is first a comparison with
categories represented in the FRMS. Here, if a category is represented in the FRMS, the present model predicts that exemplars need not be in the LCPM for a new item to be associated with that category. To test this assumption, imagine two lists of 30 items. In the first list condition, these 30 items are three members from each of ten categories. The first two items are from the first category as is the twenty-first item; the third and fourth items are from the second category as is the twenty-second item, etc. The second list condition is the same as the first-list condition except that the first ten items in odd-numbered presentation positions have been replaced by ten new, unrelated items. Recall of the last ten items, which are the same for both conditions, is the dependent variable. The present model would predict better recall of the last ten items for the first-list condition, but a model which contends that all comparison is done in the LCPM (e.g., Kintsch, 1970) would predict no differences between the conditions in recall of these items.

The third line of research involves the differential availability of markers. No particular experiments are suggested, but an interesting research programme might centre around identifying various types of markers, seeing which types are more available as a function of the type of stimuli, and testing the effects of various manipulations in changing the availability of markers.

Recognition and Recall. The description of recognition differences as a function of homographic encoding can be tested. The typical paradigm of spaced versus adjacent presentation of two-item categories can be used.
Two additional groups would be presented the same spaced and adjacent lists; however, category names would be provided alongside the items for the two groups. This manipulation should reduce the possibility of differential bias between the two groups. Thus, an interaction between spacing and whether the category names are presented would be expected.
REFERENCES


Guttman, L. A general nonmetric technique for finding the smallest coordinate space for a configuration of points. *Psychometrika*, 1968, 33, 469-506.


Miller, G.A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 1956, 63, 81-97.


Schwartz, R.M., & Humphreys, M.S. List differentiation in part-whole free recall. *American Journal of Psychology*, 1972, in press. (b)


Tulving, E. Subjective organization in free recall of 'unrelated' words. *Psychological Review*, 1962, 69, 344-354. (b)


