The use of this thesis is subject to the usual restrictions that govern the use of manuscript material. Reproduction or quotation of the text is permitted only upon written authorization from the author of the thesis and from the academic department by which it was accepted. Proper acknowledgment must be given in all printed references or quotations.
EFFECTS OF AGE AND CUING ON RETRIEVAL FROM SEMANTIC MEMORY

by

Raymond William Horn

Department of Psychology
Duke University

Date: _______________________

Approved:

Herbert F. Crovitz, Supervisor

Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Psychology in the Graduate School of Duke University

1979
ABSTRACT
(Psychology-Clinical)

EFFECTS OF AGE AND CUING ON RETRIEVAL FROM SEMANTIC MEMORY

by
Raymond William Horn
Department of Psychology
Duke University

Date: ___________________________
Approved: ______________________

Herbert F. Crovitz, Supervisor

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Psychology in the Graduate School of Duke University

1979
ABSTRACT

EFFECTS OF AGE AND CUING ON RETRIEVAL FROM SEMANTIC MEMORY

by

Raymond William Horn

Elderly subjects are known to perform less well than young subjects on laboratory tests of recall from episodic memory. Although the elderly report increased difficulty in recalling information from semantic memory, experimental attempts to demonstrate this deficit are equivocal.

It is suggested that studies which use multiple choice tests to measure recall from semantic memory fail to find age-related deficits because the tests provide cues to aid in recall, a procedure known to reduce age-related differences in recall from episodic memory. When time to retrieve a single item of information from semantic memory is measured, some studies show an age-related deficit while others do not.

When episodic recall is tested using categorized lists, the elderly show recall deficits largely because they access fewer categories than do young subjects. Semantic cues increase the number of categories recalled by the elderly subjects more than for young subjects in such tasks.
Since studies with young subjects show that recall both from categorized lists and from a taxonomic category (a semantic recall task) proceeds via temporal clusters of related items, it was hypothesized that elderly subjects would show increased difficulty in accessing clusters of related items in a semantic recall task, just as they do in recall of categorized lists. Further, it was hypothesized that semantic cues would reduce the time taken by the elderly to access sequential clusters of information from semantic memory.

In one experiment, healthy, well-educated young (ages 19-21) and old (ages 67-72) subjects were required to perform a Bousfield task: to generate examples from two taxonomic categories, foods and animals, for 15 minutes. The slope-difference algorithm, a procedure developed by Gruenewald and Lockhead, was used to categorize each subject's inter-item times (IIT's) into times between temporal clusters (BIIT's) and times between items within temporal clusters (WIIT's). In a second experiment, a group of old subjects were given semantic differential labels as cues for recall on one of their two experimental trials.

Results for the first experiment showed no age effect on mean BIIT, number of clusters, or average cluster size for recall of food items. There were also no age effects during the first 5 minutes of recall of animals. Later in the task old subjects had longer mean BIIT's for animals than did young subjects. The differences appeared to result because old subjects tended to report primarily mammals, while young subjects
reported birds, fish, reptiles/amphibians, and insects as well. A trend toward slower mean WIIT's for old subjects was attributed to slower vocalization rates. Thus, Experiment 1 failed to demonstrate age-related differences in time to access successive clusters of related items in semantic memory or in the rate at which items in a cluster are emitted. Higher repetition rates observed for the old subjects do support an age-related deficit in recognition.

In the second experiment, only half the subjects reported that the semantic-differential cues were helpful in finding new items. No effect of cuing was observed for the food category. Cuing did significantly reduce mean BIIT for animals during the last 5 minutes of recall. However, the actual effect of cuing on number of clusters produced was minimal. It was suggested that more practice with the cues might have led to higher cue usage and a greater impact on BIIT.
This work is dedicated,
with love and gratitude,
to my parents.

R. W. H.
ACKNOWLEDGMENT

Although many individuals have provided aid, guidance, and inspiration at important times during the work on this dissertation, Dr. Herb Crovitz has provided support and encouragement from the beginning to the end. Herb has guided my work by example, not by dictum. His comments have been helpful and to the point, although often I could not see their utility at the time. In the final analysis, his "hit rate" has been very good. From him I will take away memories of stimulating talks and two pieces of advice, "Think small" and "Whenever you can, count."

My other committee members have each been helpful with specific problems and questions, but I believe their most important impact on me has come from who they are. Dr. Greg Lockhead has consistently shown me that science can be done for fun, and done well. Dr. Irwin Kremen's passion and single-minded pursuit of truth in expression has provided me an example to emulate. It was because of Krem that I came to Duke, and I will sorely miss him when I leave. Dr. Pat Logue reminds me that a clinician can also be a scientist, and Dr. John Artley shows me, by his example, that a scientist can be gentle and have a twinkle in his eye.

Without the help of Paul Gruenewald this dissertation literally could
not have been completed. He developed the measurement paradigm which I have used, and has provided computer programming when I needed it.

I would particularly like to acknowledge the assistance of Dave Sanders, whose dedication to precision in measurement gave this dissertation a solid quantitative base. Because of his work, I can believe in my numbers.

Although separated from me in space, Mick Smyer has supported and encouraged me as only a friend can do. I owe more to him than I can express.

Particular gratitude must go to Edna Bissette, who typed this manuscript. She was good-natured and tolerant, even when it seemed like English must be my second language. Errors which remain are in spite of her, not because of her.

Finally, I would like to thank Jane Hewitt for her help and support at critical points, and for her toleration of the long periods between "temporal bursts." She has helped me look at myself and discover the confidence to make my own choices, live with my own mistakes, and finally finish.

This dissertation was supported in part by grant 90-A-1048 from the Administration on Aging, Department of HEW. Researchers undertaking such projects are encouraged to express freely their professional judgment. Therefore, points of view or opinions stated in this document do not necessarily represent the official position or policy of the Department of HEW.
CONTENTS

ABSTRACT iii

ACKNOWLEDGMENT vii

LIST OF TABLES xi

INTRODUCTION 1

Age-Related Deficits in Retrieval from Semantic Memory 2
Age-Related Deficits in Retrieval from Episodic Memory 7
Time as an Indicator of Difficulty in Retrieval 10
Measuring the Time Course of Semantic Free Recall 11
Measuring the Time Course of Episodic Free Recall 17
Summary and Overview of the Experiments in the Present Study 19

METHOD 23

Subjects 23
Procedure 24
Design 26
   Experiment 1 26
   Experiment 2 27
Data Analysis 27

RESULTS 31

Numerical Replication of the Slope-Difference Algorithm 31
Experiment 1 31
   Number of Words 31
   BIIT's 35
   WIIT's 40
   Number of Clusters and Cluster Size 42
Subcategories of Animals 43
Repetitions 45
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Comparison of Young Subject Data for Mean BIIT's and Mean WIIT's by Category and Interval: Gruenewald and Lockhead vs. the Present Experiment</td>
<td>32</td>
</tr>
<tr>
<td>2. Comparison of Young Subject Data for Mean Cluster Size and Mean Number of Clusters by Category and Interval: Gruenewald and Lockhead vs. the Present Experiment</td>
<td>33</td>
</tr>
<tr>
<td>3. Mean Number of Words Produced by Age, Category, and Interval</td>
<td>34</td>
</tr>
<tr>
<td>4. Mean BIIT in Seconds by Age, Category, and Interval</td>
<td>36</td>
</tr>
<tr>
<td>5. Mean WIIT in Seconds by Age, Category, and Interval</td>
<td>41</td>
</tr>
<tr>
<td>6. Mean Total Number of Clusters and Mean Number of Single and Multiple Item Clusters by Age and Category</td>
<td>41</td>
</tr>
<tr>
<td>7. Mean Percent of Clusters Which Are Single Item Clusters by Age and Category</td>
<td>44</td>
</tr>
<tr>
<td>8. Mean Size of Multiple Item Clusters by Age and Category</td>
<td>44</td>
</tr>
<tr>
<td>9. Number of Animals Recalled by Subcategory for Young and Old Subjects Used in Experiment 1</td>
<td>46</td>
</tr>
<tr>
<td>10. Mean Percent Repetitions by Age and Category</td>
<td>47</td>
</tr>
<tr>
<td>11. Mean BIIT in Seconds by Category, Cuing Condition, and Interval for Elderly Subjects</td>
<td>50</td>
</tr>
</tbody>
</table>
12. Mean WIIT in Seconds by Interval for Elderly Subjects
13. Mean Number of Clusters by Category, Cuing, and Interval for Elderly Subjects
14. Mean Size of Multiple Item Clusters by Category and Cuing Condition for Elderly Subjects
INTRODUCTION

In 1972 Tulving introduced the concepts of "semantic" and "episodic" memory (Tulving, 1972). The semantic memory system is considered to involve memory for language, rules for the manipulation of words and symbols, patterns of behavior, and all other information which does not involve a temporal identifying feature. The episodic system is thought to receive and store information about dated personal events or episodes and the temporal-spatial relations among events. Using these definitions, the meaning of a word or an item of general information (e.g. the boiling point of water) is stored in and retrieved from semantic memory. What an individual had for breakfast, where he spent last Sunday or what words were on the list he just learned in a psychology experiment are all items of information stored in and retrieved from episodic memory.

In the following pages we will use Tulving's semantic/episodic distinction to organize discussions of age-related deficits in memory retrieval and of the effect of cuing on those deficits. Then we will review the results of studies of the time course of both episodic and semantic free recall tasks in young subjects in order to present a method for studying the temporal properties of recall difficulties in the elderly. Finally, two
experiments will be presented which examined the effect of age and cuing on the temporal properties of semantic free recall.

**Age-Related Deficits in Retrieval from Semantic Memory**

Although older individuals frequently complain of difficulty in recalling familiar or well-learned information (Lowenthal & Berkman, 1967), laboratory attempts to demonstrate this retrieval difficulty from semantic memory have produced mixed results, with some studies finding no age-related deficits and others finding moderate deficits. For example, vocabulary test scores have been found to decrease only slightly with age (Botwinick & Storandt, 1974). Also, when old and young subjects were asked to produce exemplars of taxonomic categories (e.g. animals, trees, fruits) for one minute, Drachman and Leavitt (1972) found no age-related difference in the number of items produced.

Although the details of the study have not yet been published, Fozard and Poon (1978) reported finding no age differences in number of correct responses on a current events questionnaire. These authors also reported that "retention of colloquialisms for subjects over 50 years of age was relatively intact" (p. 7) as measured by the same instrument.

The conclusions of Fozard and Poon appear at odds with those of Warrington and Silberstein (1970) and Warrington and Sanders (1971). These latter authors also administered questionnaires about current events and also asked subjects to name the pictures of famous persons who had been in the news over the last 50 years. While all age groups showed
higher performance for more recent material, older subjects did less well than the younger groups. However, much of the recall difference between young and old subjects was eliminated when a multiple-choice version of the test was employed. The questionnaire used by Fozard and Poon was also a multiple-choice instrument. Thus the apparently discrepant findings of these studies could be reconciled by assuming that older subjects had as much information "available" in semantic memory as young subjects, but that the information was less "accessible" (Tulving & Pearlstone, 1966). The multiple-choice format can be thought of as providing cues which help overcome a deficit in retrieval from semantic memory.

In at least some of the above studies, cuing facilitates recall from semantic memory while free recall shows an aging deficit. For reasons which will be elaborated below, cuing can be thought of as facilitating recall by decreasing the time taken to access information in memory. Thus, the question arises as to whether there exist age-related deficits in retrieval times from semantic memory.

All of the published studies of retrieval rates from semantic memory involve measuring latencies to retrieve a single item (e.g. a word), or a single cluster of items (e.g. six letters). Eysenck (1975) gave young and old subjects the task of producing a word from a particular taxonomic category beginning with a particular letter (for example, "Name a fruit which begins with 'A'.") He found no difference in response times between young and old subjects, and concluded that rates of search in semantic
memory were the same.

However, Thomas, Fozard, and Waugh (1977) found that the latencies to name pictures of familiar objects increased with age. In one condition of this experiment, the need to retrieve the name of the picture from semantic memory was eliminated by first presenting the name of an object, then presenting a picture, and finally requiring subjects to name the picture. On half the trials the picture was the same as the presented word and on half it was different. When the name and picture matched, the age difference in naming latency disappeared.

In a second experiment (Thomas, Waugh, and Fozard, 1978), these authors used a modified Sternberg paradigm (Sternberg, 1966) to measure the rate at which subjects retrieved information stored in semantic memory. The subjects' task was to answer yes or no to whether a probe letter was a member of a previously learned list of letters. A "familiar list" (a, b, c, d, e, f) and an "unfamiliar list" (p, g, k, t, r, i) were used. Presumably the very common sequence of the first six letters of the alphabet is stored in semantic memory, while the unfamiliar set (since it was newly learned for the experiment) was stored in episodic memory. Results indicated that, for both familiar and unfamiliar lists, older subjects took longer to determine whether a probe was in the target list than did young subjects, and that the effect was more pronounced for the unfamiliar set (100 msec for the familiar set vs. 350 msec for the unfamiliar set). In terms of the semantic/episodic distinction, we would conclude that, in the
elderly, there is some slowing in retrieval of information from semantic memory, and a larger slowing of retrieval of information from episodic memory.

Since vocabulary tests are typically untimed (Wechsler, 1955), the indication of slower retrieval from semantic memory in the elderly is not inconsistent with a lack of age differences in vocabulary scores. Scores are based on what is recalled, not on how long recall takes. Also, since the age differences, when they are observed, in the retrieval of single items or single clusters of items of information from semantic memory are very small (100 msec in the Thomas, Waugh, and Fozard study), it is possible that Drachman and Leavitt's dependent measure, which indicated no difference in the number of words recalled from semantic memory in one minute, was not sensitive enough to detect age differences in retrieval times. If such differences were to be found, their detection would require measuring the times between individual items and determining whether those times were longer for old vs. young subjects.

While young and old subjects may differ on mean retrieval times from semantic memory, Thomas, Waugh, and Fozard (1978) also present data which show that old subjects exhibit more variability in retrieval times, both among a given subject's repeated measures and between subjects in a given age group. These authors present the "intra-subject variability" (the group mean of the standard deviation of each subject's correct response latencies) and the "extra-subject variability" (the standard
deviation of the means of the subjects' correct latencies) for all age groups in their study. In general, there is an increase in the variability of the average subject's latencies with increasing age. Also, there is an increase in the variability between subjects within a given age group with increasing age. The authors conclude that "such [age-related] differences might indicate less consistency in the strategies used by older subjects" (p. 532). They also note that, while the fastest times of older and younger subjects were similar, the longest times were more extreme for older subjects, i.e. the elderly distributions were more skewed. In a separate examination of these skewed distributions, Fozard and Poon (1978) state that "this finding [of less consistency of retrieval times in the elderly] could not be accounted for by the existence of a positive correlation between longer response latencies and variability of response" (p. 10).

The above conclusions on increased variability in memory search for elderly subjects point out two difficulties which may be encountered in attempting to measure age differences in memory retrieval rates. First, greater within-group variability decreases the chance of finding statistically significant mean group differences. Older subjects as a group tend to be more heterogeneous in performance than a corresponding group of young subjects. (This finding is true for performance on many different tests of memory and cognitive functioning, as indicated by Botwinick & Storandt, 1974.) Also, the variability may tend to obscure the fact that, while some of the older subjects in a group may perform as well as younger
subjects, mean differences may result from a few older subjects who perform much less well. Second, if old subjects are more variable as individuals than young subjects in how they respond to a given experimental task, a conclusion of no mean differences between groups could obscure the fact that the age difference may show itself as a difference of variability and not of means.

**Age-Related Deficits in Retrieval from Episodic Memory**

Although the studies by Warrington and Silberstein (1970), Warrington and Sanders (1971), and Thomas, Fozard, and Waugh (1977) indicate that age-related deficits exist in retrieval from semantic memory, and that those deficits can be reduced with cuing, these effects are more pronounced for the retrieval of information from episodic memory.

Laurence (1967a) tested young and old subjects on the free recall of a list composed of items from one taxonomic category (animals) versus a list composed of items from 12 different taxonomic categories. She discovered minimal age differences on the former but significant age differences for the latter. In a follow-up study, Laurence (1967b) found that presenting the names of the 12 categories at recall eliminated the age difference in amount recalled.

When Drachman and Leavitt (1972) provided subjects with first-letter cues for items on learned lists, they found no reduction in the young/old recall difference. However, when Smith (1977) provided the names of the
taxonomic categories containing the words as cues to recall, the age difference was eliminated.

Craik (1977) suggests that these studies, which show that cuing improves episodic recall more for older than for younger subjects, indicate that older subjects benefit more than young subjects from information which guides or delimits their retrieval process. He suggests that, as the domain from which a list is constructed becomes larger (e.g. the domain from which a list of animals is chosen has more members than does the domain from which a list of one-place digits is chosen), the elderly will experience greater difficulty in recalling the particular list items. Craik supports this position with data published in 1968 (Craik, 1968). In that experiment he constructed lists from domains of increasing size (one-place digits, English county names, animals, and unrelated words). Within each domain, a series of lists of increasing length was developed so that both list length and domain size could be varied independently. In general, as list length increased so did the disparity in recall scores between young and old subjects. However, the divergence of young and old recall totals with increasing list length was more pronounced as the domain from which the list was constructed became larger.

The impact of cues presented at retrieval has been explored using categorized lists also. In this type of experiment, lists typically are composed of 3 to 5 exemplars of each of 4 to 10 different taxonomic categories. For example, a 16-item list may contain 4 animals, 4 trees, 4 men's
names, and 4 fruits. A category is scored as recalled if a subject recalls at least one item from it.

Using such categorized lists, Hultsch (1975) found that older subjects recalled both fewer categories and fewer items per category under non-cued conditions than did young subjects. However, when the category names were presented at recall, older subjects recalled as many categories and as many items per category as did uncued young subjects. Although cuing also increased the number of categories recalled by young subjects, the percentage increase in number of categories recalled was larger for the older subjects (76% vs. 35%). Smith (1977) has also shown a greater percentage increase in categories recalled for older subjects relative to younger subjects when category cues are presented at recall.

The effectiveness of cues in increasing recall in the Smith and Hultsch experiments suggests that the older subjects had learned the list information, but were temporarily unable to recall it. This interpretation is consistent with findings by Buschke (1974). Buschke presented young and old subjects with a list of unrelated words followed by a series of recall trials. The list was not readministered between the trials. He observed that older subjects showed more variability than young subjects in which particular words were recalled on successive trials. Once the young subjects recalled a word, it tended to be recalled on all subsequent trials. The old subjects, however, produced more instances of recalling a word, failing to recall it on the next trial, and then subsequently recalling it on a later
trial. Buschke interpreted these results as indicating that the older subjects had learned the words, but showed more difficulty in consistently accessing them during retrieval. Since the recall of categorized lists typically proceeds by the recall of items from one category at a time (see e.g. Tulving & Pearlstone, 1966), the Hultsch and Smith data suggest that, at least with categorized lists, the elderly subjects have more difficulty accessing successive "clusters" of information, i.e. the categories themselves. Cuing helps to overcome this difficulty in access.

**Time as an Indicator of Difficulty in Retrieval**

In the discussion of the above studies, we have frequently used the expression "difficulty in retrieval." This phrase has three different meanings depending on the dependent variable used in the studies. "Difficulty" can manifest itself as less information recalled (e.g. Drachman & Leavitt, 1972; Hultsch, 1975), as a longer time to retrieve information that is recalled (e.g. Thomas, Fozard, & Waugh, 1977), or as increased variability in retrieval times (Thomas, Waugh, & Fozard, 1978). Thus, even in studies which show no age difference in amount recalled (e.g. Drachman & Leavitt, 1972), age differences may exist in retrieval times or in the variability of retrieval times. Only by monitoring retrieval times as well as amount recalled can we clarify the meaning of "difficulty in retrieval" for a particular recall task.

None of the studies mentioned above have monitored the time course of free recall of a series of items from either semantic or episodic
memory. Only the studies from the Fozard laboratory and the experiment by Eysenck have examined retrieval times at all, and they have done so for the recall of single items or single clusters of information like the six-letter list in the Thomas, Fozard, and Waugh study. We have found no studies in the literature which have compared times taken to recall the items in a free recall experiment for young and old subjects.

Studies have been made of the time course of recall for younger subjects using the recall of exemplars of a taxonomic category (e.g. Bousfield, 1944; Dean, 1971; Gruenewald & Lockhead, Note 1) and the recall of categorized lists (Patterson, Metzler, & Mandler, 1971; Pollio, Richards, & Lucas, 1969) as experimental tasks. Thus methods exist which can be used to study the temporal properties of free recall in elderly subjects. Since free recall is a series of events which can extend over a period of many minutes, measuring the time structure of free recall in the elderly could indicate whether their difficulty in retrieval includes lengthened times to recall items successfully, as well as failure to recall some items at all.

**Measuring the Time Course of Semantic Free Recall**

Bousfield and his colleagues produced the first work on the temporal properties of recall from a taxonomic category (Bousfield, 1944; Bousfield & Sedgewick, 1944) using college students as subjects. Bousfield and Sedgewick observed that many of the items were emitted in temporal bursts, with longer pauses between bursts than between items within
bursts. They suggested, but did not establish quantitatively, that the items in the individual bursts were "associatively related."

The measurement procedure used in these experiments involved having the experimenter press a key which produced a cumulative record of total items emitted as a function of time. Each time the experimenter pressed the key, he would also write down the spoken word. Bousfield and Sedgewick acknowledged that the experimenter had trouble accurately performing both recording tasks, particularly in the early stages of recall when items were spoken quite rapidly. However, they concluded that the method was superior to an earlier procedure of having subjects write their own output in columns and shift to a new column every two minutes when given a signal by the experimenter. Recording times between spoken words eliminated time to write the words from the measurement and allowed the experimenter to monitor the more rapid initial output more closely.

In a similar study with college student subjects, Levin and Breznitz (1977) monitored the time to produce items in a "continuous association task" which required subjects to say any word which came to mind during a 90-second interval following presentation of a single stimulus word. The subjects' output consisted of two types of clusters, "stars" (a word series whose items were associatively related to the stimulus word) and "chains" (a word series whose items were more closely related to the immediately preceding word than to the stimulus word). The measures of associative
relatedness were taken by having the subjects themselves classify their lists into chains and/or stars. If the stimulus word were "chair," a star might be "armchair, rocking chair, wheelchair, electric chair . . ." and a chain might be "table, vase, flowers, flies, butterflies . . ." Times between items were measured with a stopwatch while listening to a tape recording of the subjects' output.

Levin and Breznitz eliminated "mixed clusters" (a word series containing some associations related to the stimulus word and some related to the preceding response word) and single item clusters from their analyses without providing a justification for doing so. Mixed and single item clusters together contained 40% of all words produced. From the remaining data they demonstrated that, for both star and chain clusters, from two to five times as many subjects had longer median times between clusters than between items within clusters.

Thus, if we view both recall of exemplars of a taxonomic category and continuous association to a single stimulus word as recall from semantic memory, we can conclude that recall from semantic memory is not regular, but is composed of temporal bursts of items. Also, although the composition of the clusters may vary depending on the task instructions (i.e. associative clusters in the recall of examples from a taxonomic category, and a combination of associative clusters and chains of successively associated items in the continuous association task), the items are not grouped randomly into temporal bursts, i.e. independent of their meaning.
Dean (1971) and Gruenewald and Lockhead (Note 1) have provided a quantitative demonstration that the temporal bursts of items in Bousfield's semantic free recall task are composed of meaningfully related words. After completing the free recall task, Dean's college student subjects were given a set of cards, each containing one of their emitted words, and were asked to sort the cards into categories. Dean then divided the inter-item times (IIT's) from the free recall task into quartiles and defined a sequence of words from the original emission as a "fast sequence" if all the IIT's for that sequence were in the fastest quartile. When he compared these fast sequences with "clusters" (a sequence of words emitted during the recall task which were subsequently sorted into the same category), he calculated that there was a .870 average probability that a word would be in a cluster, given that it was in a fast sequence. That is, "if a word is emitted in a fast burst of responding, the probability it is related (in the sense of being sorted together) to the other words in the same fast sequence is high" (p. 97). Dean also observed that the between cluster inter-item time (BIIT) increased as more clusters were emitted. The increase was relatively linear for the first 8 or 10 clusters and then became quite variable.

From these data, Dean concluded that, in such a search of semantic memory, the search is not a random search for words, but for groups of words or clusters which are meaningfully related. As the search is continued, it takes progressively longer to find new clusters.
Gruenewald and Lockhead's methodological approach to the phenomenon of clustering in the Bousfield task differed from Dean's in two ways. First, the classification of IIT's into between cluster inter-item times (BIIT's) and within cluster inter-item times (WIIT's) was accomplished by comparing the slope of the actual step-function of cumulative words vs. time with the slope of the continuous function which best fit the subject's cumulative words vs. time data. In effect, at any time \( t \), an IIT was classified as a WIIT if the slope of the real output function was greater than the slope of the theoretical output function by an arbitrarily pre-specified amount. The procedure for classifying IIT's as BIIT's or WIIT's was called the "slope-difference algorithm." This approach allowed Gruenewald and Lockhead to take into account the possibility that a WIIT was shorter in the beginning of the task when the subject was responding rapidly than it might be near the end when he was responding more slowly.

The other major difference in method between Dean and Gruenewald and Lockhead was that, while Dean's indicator of semantic clusters was based on having the individual subjects sort their output into categories, Gruenewald and Lockhead had independent judges rate whether each consecutive pair of items were in the same semantic cluster or not. On a scale from 0 to 3, where 0 indicates that two items are definitely in the same cluster and 3 indicates that they are definitely in different clusters, 86% of the IIT's classified as BIIT's by the judges (i.e. the IIT's had an average rating of 2.5 or greater) were also classified as BIIT's by the
slopes difference algorithm with its criterion slope difference set at .1. Thus, in spite of methodological differences in the semantic and temporal classifications of items into clusters, Gruenewald and Lockhead support the general findings of Dean, i.e. rapidly produced bursts of items tend to be composed of semantically related words.

In addition to agreeing that temporal bursts are composed of meaningfully related groups of words, the experiments by Dean and Gruenewald and Lockhead also agree on other features of the output. Both experiments found an average cluster size of approximately 2 with about 90% of the clusters having size less than or equal to 5. Gruenewald and Lockhead also observed the general increase in BIIT's with increasing output. The increase was linear in the early part of the task and became more variable as output proceeded. However, while both studies show that BIIT's increase with time on task, Gruenewald and Lockhead showed that WIIT's changed very little, even after 15 or 30 minutes of recall.

The above studies of the time course of semantic free recall support the following generalizations:

1. Output is composed of both rapidly emitted sequences of words (temporal bursts) and more slowly emitted items.

2. The temporal bursts correspond to clusters of meaningfully related items.

3. While BIIT's increase as more clusters are recalled, WIIT's remain nearly constant.
It appears that search of semantic memory involves two different types of search, each with different temporal properties. One is a search for clusters of related items. This search takes more and more time as more clusters are accessed. The second is a search for items within clusters. These times are regularly shorter than BIIT's and do not change greatly with time on task.

Measuring the Time Course of Episodic Free Recall

Conclusions similar to those concerning the time course of semantic recall have been reached in studies of episodic recall of categorized lists. These time-based studies have all used young subjects.

Using categorized lists, Pollio, Richards, and Lucas (1969) demonstrated that words tended to be recalled by categories, and that the BIIT's were regularly longer than WIIT's. For successive clusters recalled, the BIIT's appeared to increase in an exponential fashion when plotted vs. output position of the cluster. These authors suggested a two-step model in which categories were recalled and then items within the categories were recalled.

Patterson, Metzler, and Mandler (1971) elaborated the concept of a BIIT in a categorized free recall task by suggesting that such times involve three processes. A "category exit time" is the time spent in unsuccessful search for additional items in the current category. The "category access time" is the time spent finding a new category, while the "word access time" is the time to find a word in the new category. Their experiments
established two facts. First, providing subjects with the category names at retrieval eliminated the increase observed by Pollio et al. (1969) in BIIT for successive clusters. Second, even when the category access time had been reduced or eliminated by cuing, BIIT's were still reliably longer than WIIT's. Patterson et al. argued that these facts implied that the category exit time (plus a small word access time) accounted for the initial disparity between BIIT's and WIIT's, and that, without cues at retrieval, category access time increased with the number of categories recalled. As in Pollio et al.'s study, they found an exponentially increasing BIIT with number of clusters emitted.

Using the same paradigm but shorter lists (16 words vs. 25 in the above two studies), Kellas, Ashcraft, Johnson, and Needham (1973) lent support to the notion of a category exit time by showing that BIIT's were regularly longer when subjects were explicitly requested to give output by categories than when they were not so instructed. These authors observed a linear increase in BIIT with increasing cluster output position, and suggested that it might be a function of using shorter lists.

The above three studies all used word lists with an inherent category structure supplied by the experimenter. Apparently subjects used this information to help learn the list, producing associatively related clusters of rapidly emitted items at output. McCauley and Kellas (1974) extended the observations of temporal clustering to lists of unrelated words. They instructed subjects to learn the words by using image mediators, sentence
mediators, or a repetition strategy to group the words into units of three items. Times between clusters of items determined by the mediators were regularly longer than times within the clusters. These authors observed a linear increase in BIIT's for all instructional conditions, although the slope of the functions differed depending on the learning strategy used. They argue that the linear increase reflects a ready accessibility of the clusters used in these short lists and support a serial search model for clusters under these experimental conditions.

Summary and Overview of the Experiments in the Present Study

If we compare the results from studies which include a temporal description of recall from semantic memory with those which include a temporal description of the type of episodic recall involved in free recall of learned lists, the following similarities emerge:

1. Output is composed of rapidly emitted items and more slowly emitted items.

2. The temporal bursts of items correspond to clusters of words which share some common feature.

3. The common feature among items in a cluster varies with the type of task. It may involve associative relatedness, as determined by a given subject's pattern of associations, or it may reflect an organization of the material determined by the experimenter through categorizing the list or requiring certain types of mediation during learning.
4. BIIT's are regularly longer than WIIT's, with the former increasing with time on task while the latter increase only slightly, or remain constant.

Our discussion of recall failures in both semantic and episodic recall by the elderly raised questions about retrieval times which can be addressed with methods developed to study the time course of free recall by measuring times between sequentially emitted items. We concluded that the papers reviewed supported the hypothesis that elderly subjects retrieve items from episodic memory more slowly than do young subjects. Although old subjects may recall less information from semantic memory than young subjects, the question remains: Do the elderly retrieve items from semantic memory more slowly than young subjects?

Also, older subjects were aided in both semantic and episodic recall by cues. Patterson et al. (1971) suggested that cues function to shorten access time for new clusters, at least in episodic free recall. Thus a second question can be raised about retrieval times from semantic memory in the elderly: Do cues function to shorten times to retrieve clusters in semantic free recall?

We have chosen the method of Gruenewald and Lockhead to address these two above questions. By using their slope-difference algorithm to differentiate BIIT's and WIIT's in recall from a taxonomic category, we can measure two types of times in semantic recall, those between temporal clusters and those between items within temporal clusters.
We have chosen to use the Gruenewald and Lockhead method of dividing IIT's into BIIT's and WIIT's because it takes into account the individual rate of slowing of output for each subject. Dean's method of defining temporal bursts by choosing times from the lowest quartile of time makes it almost impossible for him not to define all the initial items as being in a fast sequence since in the first minute or two of the task nearly all the items are in the lowest quartile. However, since Gruenewald and Lockhead's method takes into account the instantaneous rate of output, it allows us to define temporal bursts in the initial minutes, even if the times between them fall in the lowest quartile of all IIT's.

Two experiments were performed using the semantic free recall task and slope-difference algorithm measuring system used by Gruenewald and Lockhead. Experiment 1 was designed to test whether old subjects retrieved clusters from semantic memory more slowly than young subjects. Young and old subjects were asked to spend 15 minutes producing as many items as possible from one of two taxonomic categories (foods or animals). These two categories were chosen because they were used by Gruenewald and Lockhead in developing their method. Pilot work indicated that old subjects would be able to recall items throughout the entire 15-minute period used in the task. Old and young subjects were compared on total words recalled, total clusters recalled, average cluster size, and mean BIIT and WIIT for each successive 5-minute interval.

Experiment 2 was designed to test whether cues decreased the time
taken by elderly subjects to retrieve clusters from semantic memory. Elderly subjects were tested on the Gruenewald and Lockhead task under cued and non-cued conditions. In order to have cues appropriate for semantic free recall and which could be used with both food and animal categories, Osgood, Suci, and Tannenbaum's (1957) semantic differential dimensions were used as cues.
METHOD

Subjects

Young subjects were all Duke University undergraduates who participated in the study as part of a course requirement. Prior to the experimental session, they were told that they would be participating in a "study of semantic memory." Seventeen subjects were actually run so that a final group of 12, equally balanced for sex, could be obtained whose vocabulary scores matched that of the older subject group. The 12 young subjects actually used in Experiment 1 had a mean age of 19.8 years and a mean vocabulary scale score of 14.4 on a short form of the Wechsler Adult Intelligence Scale vocabulary subtest (Jastak & Jastak, 1964).

Older subjects were solicited by phone from the subject pool of the Duke Center for the Study of Aging and Human Development. Requirements were that subjects be between 67 and 72 years old, that they had completed between 1 and 3 years of schooling past high school, and that they had no history of heart attack, stroke, high blood pressure, or cancer. A total of 31 subjects were run to achieve a final group of 24. One old subject was dropped from the study because she was over the age maximum, and 6 were dropped because they failed to follow instructions on some part of the
experiment. The final group of 24 old subjects contained 12 men and 12 women. All old subjects were compensated $4.00 for one hour of experimental time.

Half of the old subjects (6 men and 6 women) produced data used in Experiment 1. The mean age of this group was 70.2 years and the mean vocabulary scale score was 15.1. Vocabulary scale scores did not differ between the groups of young and old subjects used in Experiment 1, 

\[ t_{(22)} = 1.68, \ p = \text{n.s.} \]

The entire group of 24 old subjects provided data for Experiment 2. The mean age of this group was 70.0 years and the mean vocabulary scale score was 14.9.

**Procedure**

Experimental sessions were run in a sound-attenuated chamber containing a table and two chairs. A tape recorder (Sony, Model C-105) was placed on the table in front of the subjects.

Subjects were told that they would be asked to report examples verbally from general categories supplied by the experimenter (see Appendix A for Recall Task Instructions). They were then given a two-minute practice trial using "furniture" as the category, followed by two experimental trials, one involving foods and one animals. The order of presentation of categories was randomly determined with the restriction that half the subjects received food first and that equal numbers of men and women received food first. Each subject initiated the experimental trials by turning over a card and reading the category name aloud from it. During the
experimental categories, the experimenter was absent from the room.

For conditions in which subjects were cued, they were given special instructions immediately prior to the cued trial (see Appendix B for Cuing Instructions). Each subject was given a card face down on which were written the six main semantic differential dimension labels (i.e. good, bad, strong, weak, active, and passive) followed by the name of the category being cued (e.g. "good animals"). Subjects were told that when they began to have difficulty thinking of items they should use the phrases on the card to try to bring more examples to mind.

Although subjects were not told explicitly when to pick up the cue card, on the basis of pilot study it was expected that the subjects would slow down or run out of examples before the end of the 15 minutes. (Pilot subjects produced half of their words within 5-7 minutes.) Thus, the effects of cuing were expected in the latter part of the 15-minute time period.

Immediately following the second experimental trial, subjects were asked about their use of cues as a manipulation check. First they were asked whether or not they used the cues and whether the cues had been helpful. They were also asked to try to recall specifically which words were prompted by the cues. Then the subjects were asked to give three examples for each of the cue phrases to ensure that they could successfully interpret them and use them to classify items in the experimental category used. Sixty seconds were allowed to produce each set of three responses, and the total number of responses was calculated for each subject.
Following the manipulation check (or, in the case of the young subjects, following the second category) each subject was administered the Jastak short form of the WAIS vocabulary subtest. Then they were given a verbal fluency (VF) task consisting of verbalizing as many words as possible beginning with S in 60 seconds, followed by the same task for the letter M. VF was taken as the total words produced. Finally, subjects were given a list of 204 words typed in capital letters in two columns on three sheets of paper. The words were examples from the categories "toys," "parts of a house," "trees," and "clothing" in the Battig and Montague (1969) category norms. Items with a frequency of 10 or greater were used. Subjects were asked to read as many words as possible in one minute.

Design

Experiment 1

Data for Experiment 1 came from non-cued, first trial conditions. Twelve young subjects and the 12 old subjects who had a non-cued first trial were divided into four groups. Half of the subjects in each age group received the food category and half received the animal category. For each subject, mean BIIT and mean WIIT were calculated for each 5 minutes of the 15-minute experimental task. The data were analyzed as a 2 (age) x 2 (category) x 3 (time interval) design using a three-way analysis of variance with repeated measures on the time interval factor.
Experiment 2

For Experiment 2, the 24 old subjects were divided into four groups, counterbalancing for the category used (i.e. whether a given subject was cued on the food or animal category) and trial cued (i.e. whether the cued condition came on the first or second experimental trial). Thus, each subject produced examples from both the food and animal categories, was cued on one of these categories, and was cued on either the first or second category.

The 12 old subjects whose first trial data were included in Experiment 1 provided both first and second trial data for Experiment 2.

The BIIT and WIIT data were analyzed in a 2 (category cued) x 2 (trial cued) x 2 (cued vs. non-cued) x 3 (time interval) analysis of variance with repeated measures on the last two factors.

Data Analysis

A research assistant, who was blind to the experimental procedure and hypotheses, measured the IIT's from each subject's tape-recorded output. The assistant sat at a keyboard in a soundproof chamber and pressed a key each time the subject began to speak a word. The times between consecutive words were calculated by means of a PDP-8 computer and corresponded to the times between key presses.

To determine the precision of the research assistant's time measures, 12 data files were selected at random, 6 from the old subjects and 6 from the young. For each subject, 20 consecutive times were
remeasured. Four of the sets of times were taken from the beginning of the respective protocols, four from the middle, and four from the end. No differences between correlations of the times, maximum difference between times, or average difference between times were observed between young and old subjects or at different times during the protocol. The average correlation between the times was .999995 (range .999994-.999997). The average maximum difference between times was .17 seconds, while the average difference between times was .06 seconds. The single greatest difference between any two times examined was .40 seconds. Thus a conservative estimate of the precision of each measure would be $\pm .2$ seconds, with many of the values being more precise than that. If, in the original measurements of IIT's, any value was less than 1 second, the value was remeasured twice and the average of the three measurements was taken as the value for that IIT.

The resulting IIT's provided the raw data for this experiment. These data were then entered (to the nearest .06 seconds) into the curve fitting program developed by Paul Gruenewald and were fit to a hyperbolic curve using an interactive procedure minimizing the sum-squares differences between predicted and observed number of items over time. The data were fitted to a hyperbolic function, which earlier work by Lockhead and Gruenewald had shown provided a good representation of the data. The Gruenewald and Lockhead work had compared the hyperbolic with a more usually used exponential curve and found that in nearly all cases the
hyperbolic function provided a lower best fit sum of squares than did the exponential function. As a check on whether the hyperbolic curve provided the best fit for the older subjects as it did for the younger ones used in the Gruenewald and Lockhead study, eight old subjects were chosen at random and their data were also fit to the exponential curve. In all eight cases the hyperbolic fit provided a smaller sum of squares for the best fit.

Thus, at this point in the analysis we had the parameters for the hyperbolic equation \( N = \frac{at}{b+t} \) which provided the best smooth curve fit for each subject. At this point the slope-difference algorithm was applied to each subject's data.

The slope-difference algorithm examines the actual slope between each consecutive pair of words (the inverse of the IIT) and compares that slope with the theoretical slope at the second word as determined from the best fit equation. If the real slope minus the theoretical slope is greater than an arbitrarily chosen criterion, that pair of items is considered to be in a cluster. This process is repeated for all pairs of words. In effect, a pair of words are placed in a cluster if they are emitted at a rate faster than the subject's average output rate at that point in time. Thus, the slope-difference algorithm automatically adjusts the criterion for a temporally defined cluster to take into account the general slowing of word production with time on the task.

The criterion for the slope-difference algorithm for this study is the same as that used by Gruenewald and Lockhead, .1. Although they
arbitrarily decided on this criterion, my examination of their data reveal it to be a good choice. Gruenewald and Lockhead used judges' ratings of semantic clusters to validate their algorithm's ability to produce the "basic category structure" of a subject's output based on time parameters alone. Using the .1 criterion, 80% of the items that all judges agreed were within clusters were placed in clusters by the algorithm and 88% of all items indicated as between clusters were so classified by the algorithm. Raising the criterion to .15 yields percent agreements of 68% and 92%, respectively, while lowering the criterion to .05 yields percentages of 90% and 69%, respectively. Thus, a value of .1 does appear to optimize the match between judges' ratings of semantic clusters and the algorithm's rating of temporal clusters.

Following Gruenewald and Lockhead's method, the data on BIIT's and WIIT's were divided into three intervals corresponding to successive five-minute intervals of the task. For each of the intervals an average BIIT and an average WIIT were calculated for each subject. These data were analyzed by an analysis of variance as indicated above.
RESULTS

Numerical Replication of the Slope-Difference Algorithm

Tables 1 and 2 compare the data derived from the slope-difference algorithm in the Gruenewald and Lockhead study with those obtained for the young subjects used in the present study for both the food and animal categories. The means and standard deviations for BIIT and WIIT for each time interval are found in Table 1, while means and standard deviations for cluster size (i.e. the number of words divided by the number of clusters as indicated by the algorithm) are found in Table 2. Examination of these data reveals that the numerical values obtained in the two studies are in close agreement.

Experiment 1

Number of Words

Table 3 shows the means and standard deviations for number of words produced in each time interval by age and category. Older subjects produced about as many food items as did young subjects in each interval. However, old subjects produced fewer animal items in each interval than did the young. A $2 \times 2 \times 3$ factor analysis of variance of these data revealed only significant main effects for category, $F(1, 20) = 9.27$, 

31
Table 1
Comparison of Young Subject Data for Mean BIIT's and Mean WIIT's by Category and Interval:
Gruenewald and Lockhead vs. the Present Experiment

<table>
<thead>
<tr>
<th>Source</th>
<th>Category</th>
<th>Mean BIIT in Seconds</th>
<th>Mean WIIT in Seconds</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>I</td>
<td>II</td>
</tr>
<tr>
<td>G. &amp; L. (n = 4)</td>
<td>Food</td>
<td>M</td>
<td>6.10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SD</td>
<td>(1.10)</td>
</tr>
<tr>
<td>Horn (n = 6)</td>
<td>Food</td>
<td>M</td>
<td>5.64</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SD</td>
<td>(1.42)</td>
</tr>
<tr>
<td>G. &amp; L. (n = 8)</td>
<td>Animals</td>
<td>M</td>
<td>9.68</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SD</td>
<td>(2.48)</td>
</tr>
<tr>
<td>Horn (n = 6)</td>
<td>Animals</td>
<td>M</td>
<td>9.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SD</td>
<td>(3.15)</td>
</tr>
</tbody>
</table>
Table 2
Comparison of Young Subject Data for Mean Cluster Size and Mean Number of Clusters by Category and Interval: Gruenewald and Lockhead vs. the Present Experiment

<table>
<thead>
<tr>
<th>Source</th>
<th>Category</th>
<th>Mean Cluster Size</th>
<th>Mean Number of Clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>I</td>
<td>II</td>
</tr>
<tr>
<td>G. &amp; L. (n = 4)</td>
<td>Food</td>
<td>M</td>
<td>1.97</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SD</td>
<td>(.14)</td>
</tr>
<tr>
<td>Horn (n = 6)</td>
<td>Food</td>
<td>M</td>
<td>1.77</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SD</td>
<td>(.26)</td>
</tr>
<tr>
<td>G. &amp; L. (n = 8)</td>
<td>Animals</td>
<td>M</td>
<td>1.90</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SD</td>
<td>(.18)</td>
</tr>
<tr>
<td>Horn (n = 6)</td>
<td>Animals</td>
<td>M</td>
<td>1.92</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SD</td>
<td>(.28)</td>
</tr>
</tbody>
</table>
Table 3

Mean Number of Words Produced by Age, Category, and Interval

<table>
<thead>
<tr>
<th></th>
<th>Young</th>
<th></th>
<th></th>
<th>Old</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Time Interval</td>
<td></td>
<td></td>
<td>Time Interval</td>
<td></td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>II</td>
<td>III</td>
<td>N_{Tot}</td>
<td>I</td>
</tr>
<tr>
<td>Food</td>
<td>M</td>
<td>82.23</td>
<td>53.83</td>
<td>36.17</td>
<td>172.23</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>(21.17)</td>
<td>(18.32)</td>
<td>(17.86)</td>
<td>(53.53)</td>
</tr>
<tr>
<td>Animals</td>
<td>M</td>
<td>58.67</td>
<td>35.33</td>
<td>25.67</td>
<td>119.67</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>(18.89)</td>
<td>(22.07)</td>
<td>(19.65)</td>
<td>(59.29)</td>
</tr>
</tbody>
</table>
\[ p < .01, \text{ and time interval, } F (2, 40) = 60.08, \ p < .001. \] No other main effects or interactions were significant. Thus, with the sample size used in the experiment (\( n = 6 \)), the apparent age effect on number of animals produced did not achieve statistical significance.

**BIIT's**

Table 4 shows the means and standard deviations for BIIT in each time interval by age and category. The data generally show an increase in BIIT with increasing time interval for both young and old subjects for both categories. The BIIT's for foods are nearly identical in corresponding intervals for the two age groups. However, while the BIIT's for animals in Interval I are the same for the two age groups, the values in the last two intervals show old subjects' BIIT's becoming longer more rapidly than do young subjects'. Also, BIIT's for animals are consistently longer than for foods under all conditions.

The analysis of variance of BIIT data revealed a significant main effect for category, \( F (1, 20) = 5.90, \ p < .02, \) and time interval, \( F (2, 40) = 4.50, \ p < .02. \) None of the other main effects or interactions were significant. In particular, the age by category by interval interaction, which would reflect the apparently greater rate of increase of BIIT for old subjects relative to young on the animal category, did not reach significance, \( F (2, 40) = 1.21, \ p < .31. \) The small \( n \)'s used in Experiment 1 and the large variances for old subjects in the last two intervals for animals may have been responsible for the lack of a significant three-way interaction.
Table 4
Mean BIIT in Seconds by Age, Category, and Interval

<table>
<thead>
<tr>
<th>Category</th>
<th>Young</th>
<th>Old</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Time Interval</td>
<td>Time Interval</td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>II</td>
</tr>
<tr>
<td>Food</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>5.64</td>
<td>8.92</td>
</tr>
<tr>
<td>SD</td>
<td>(1.42)</td>
<td>(2.61)</td>
</tr>
<tr>
<td>Animals</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>9.07</td>
<td>16.42</td>
</tr>
<tr>
<td>SD</td>
<td>(3.15)</td>
<td>(8.38)</td>
</tr>
</tbody>
</table>
Since the cell variances for BIIT in Experiment 1 appear related to both the cell mean and level of the interval factor, a log transform was done on the data to reduce the variance (Winer, 1962). The resulting analysis of variance yielded the same pattern of results as were obtained in the first analysis, i.e. the category and time interval main effects were the only ones significant.

Since the mean BIIT values in Table 4 strongly suggest that young and old subjects responded differently to the animal category in the last two time intervals, a post hoc, repeated measures design was devised in an attempt to reduce error variance as much as possible in the analysis of variance.

In the original design for the analysis of Experiments 1 and 2, some of the old subjects provided data for both experiments. Using independent groups of old subjects in the two experiments would have required a total of 36 subjects, a number larger than could be obtained from the Aging Center subject pool and still meet the strict criteria established for age range, educational level, and health status. Even if additional subjects could have been obtained, the data preparation and analysis alone would have required at least an additional 8 to 10 hours per subject.

Since each old subject was to be cued in one of his conditions, only by using data from the first trial was it possible to compare young and old subjects under identical experimental conditions. However, using data from only one trial for each subject in Experiment 1 prevented using a
repeated measures design, which would have allowed statistical control for between-subject differences.

As we will see in the discussion of Experiment 2, there were no observable effects of cuing on the food category for the older subjects, although there were effects for the animal category. Thus, in order to perform a more sensitive analysis of the apparent age by category by interval effect seen in Table 4, a repeated measures analysis of variance was performed on the BIIT data using both trials for each young subject along with data from old subjects who received the cuing manipulation on the food category. That is, the data were treated as if the old subjects had not been cued on the food category. Clearly, we cannot know for certain the effect of such an assumption on the resulting analysis. Thus, any interpretation of the analysis which follows must be considered as suggestive, not conclusive. It was performed only to examine the effect of using the repeated measures strategy on our ability to test the statistical significance of an apparent age by category by interval interaction.

When the data for the second trial for young subjects were analyzed, one subject's IIT data could not be successfully fitted to a hyperbola, i.e. the curve-fitting program was unable to determine an asymptote for the function. This subject (a male), who produced 192 food words, did not slow his rate of item production enough in 15 minutes for the data to be described by a hyperbolic function. Thus, both trials for this subject were eliminated from the analysis. To achieve equal n's for use in the analysis
of variance program, one older subject was also eliminated. It was arbitrarily decided to eliminate that old subject whose total words for each category was closest to the totals for the young subject. The young subject produced 192 foods and 130 animals. An older subject (a male) who produced 191 foods and 131 animals was also eliminated from the analysis.

After collapsing over trials, a 2 (age) x 2 (category) x 3 (interval) analysis of variance with repeated measures on the last two factors was performed. Significant main effects for age, \( F(1, 20) = 4.65, p < .05 \), category, \( F(1, 20) = 10.21, p < .01 \), and interval, \( F(2, 40) = 8.58, p < .001 \), were qualified by a two-way interaction of age by category, \( F(1, 20) = 7.68, p < .02 \), and a three-way interaction of age by category by time interval, \( F(2, 40) = 3.44, p < .04 \). Thus, given the qualification of this analysis mentioned above (i.e. the analysis was post hoc and compared data collected under somewhat different experimental conditions), it nonetheless did support the apparent three-way interaction observed for mean BIIT's displayed in Table 4.

Inspection of the standard deviations in Table 4 also indicates that, except for Interval I for the animal category, the old subjects show apparent greater within-group variability (as indicated by the standard deviations) than do young subjects. Since each subject's mean BIIT for an interval was based on a different number of IIT's, it was not possible to examine age-related effects on intra-subject variability as was done in the study by Thomas, Waugh, and Fozard (1978).
WIIT's

The WIIT's as a function of age, category, and time interval are displayed in Table 5. (Since two old subjects produced no multiple item animal clusters in Intervals II and III, means are based on n = 4.) Examination of the means reveals an increase in WIIT in all conditions between Intervals I and II, with a leveling off between Intervals II and III. Also, for all time intervals, the WIIT's appear shorter for young subjects than for old subjects, regardless of category.

An analysis of variance of the WIIT data confirmed a significant main effect for interval, $F(2, 40) = 13.41, p < .001$. The mean effect for age did not achieve significance, $F(1, 20) = 3.28, p < .10$, although a trend was indicated. Neither the category main effect nor any of the interactions approached significance.

Because the slope-difference algorithm increases the criterion for an IIT to be classified as a WIIT as a function of time on task, we would expect there to be some increase in WIIT simply due to the properties of the algorithm. For an average subject, the maximum IIT allowed to be classified as a WIIT increases by about a factor of 2 over the 15 minutes of the task. Thus, the fact that some increase in the WIIT's was observed was not unexpected. However, the WIIT's did not increase as much as allowed by the algorithm.

The trend toward an overall age effect on WIIT's may reflect age differences in vocalization times as opposed to differences in memory
Table 5
Mean WIIT in Seconds by Age, Category, and Interval

<table>
<thead>
<tr>
<th></th>
<th>Young</th>
<th>Old</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Time Interval</td>
<td>Time Interval</td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>II</td>
</tr>
<tr>
<td>Food</td>
<td>M</td>
<td>1.39</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>(.15)</td>
</tr>
<tr>
<td>Animals</td>
<td>M</td>
<td>1.57</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>(.22)</td>
</tr>
</tbody>
</table>

\(a\) \(n = 4\).

Table 6
Mean Total Number of Clusters and Mean Number of Single and Multiple Item Clusters by Age and Category

<table>
<thead>
<tr>
<th></th>
<th>Total Clusters</th>
<th>Single Item Clusters</th>
<th>Multiple Item Clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Young</td>
<td>Old</td>
<td>Young</td>
</tr>
<tr>
<td>Food</td>
<td>M</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>101.7</td>
<td>98.2</td>
<td>61.2</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(28.7)</td>
<td>(34.7)</td>
<td>(20.1)</td>
</tr>
<tr>
<td>Animals</td>
<td>M</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>67.2</td>
<td>48.0</td>
<td>38.7</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(32.3)</td>
<td>(19.6)</td>
<td>(15.9)</td>
</tr>
</tbody>
</table>
search rates. The average reading rate of young subjects was 19% faster than for old subjects. Overall, the old subjects' WIIT's were 19% slower than for young subjects. Regardless of the source of the difference in WIIT's, the effect is small compared to the age-related difference in BIIT's for animals in Intervals II and III.

**Number of Clusters and Cluster Size**

Hultsch's (1975) work on recall of categorized lists showed that, for episodic recall, older subjects recalled fewer categories with fewer items per category than younger subjects. In the semantic free recall task used in this experiment, we also examined total number of clusters produced and average size of cluster for the two age groups. Table 6 shows the mean total number of clusters for each category for the two age groups, as well as the breakdown of that total into single and multiple item clusters. (A multiple item cluster contains at least two items.) Although young and old subjects produced about the same number of food clusters, the old subjects produced only 71% as many animal clusters as did the young. However, only the main effect for category achieved significance by the analysis of variance, $F(1, 20) = 12.18, p < .001$. The interaction of age and category was not significant, $F(1, 20) = .450, p = n.s.$

However, when we examine the breakdown of clusters into size 1 and size 2 or greater, we observe that, although older subjects produced nearly as many single item clusters as young for both categories, and nearly as many multiple item clusters for foods, they produced only half
as many multiple item clusters for the animal category as did young subjects.

Alternatively, we can say that when older subjects produced clusters, a higher percentage of them were single item clusters. Table 7 shows the mean percent of clusters which are single item clusters as a function of age and category. An analysis of variance revealed a significant main effect for age, $F(1, 20) = 5.14, p < .05$, with no significant main effect for category or for the interaction of category by age. A Tukey test for unconfounded means revealed that no single pair of means was different at the .05 level, although the mean difference of 10.8 between young and old subjects approached the critical value of 12.10 required for significance.

Although older subjects appear to have produced fewer multiple item clusters for the animal category, when they did produce clusters, the average cluster size was the same as for young subjects. Table 8 shows means and standard deviations for average cluster size by age and category for all clusters of size greater than or equal to 2. An analysis of variance revealed no differences among these means.

Subcategories of Animals

We have seen that older subjects produced fewer items from the animal category primarily as a function of producing fewer multiple item clusters, i.e. they "ran out" of animal clusters more rapidly than did the young subjects. A possible explanation of this phenomenon is that the
Table 7
Mean Percent of Clusters Which Are Single Item Clusters by Age and Category

<table>
<thead>
<tr>
<th></th>
<th>Young</th>
<th>Old</th>
</tr>
</thead>
<tbody>
<tr>
<td>Food</td>
<td>M</td>
<td>60.30</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>(8.24)</td>
</tr>
<tr>
<td>Animals</td>
<td>M</td>
<td>58.78</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>(6.99)</td>
</tr>
</tbody>
</table>

Table 8
Mean Size of Multiple Item Clusters by Age and Category

<table>
<thead>
<tr>
<th></th>
<th>Young</th>
<th>Old</th>
</tr>
</thead>
<tbody>
<tr>
<td>Food</td>
<td>M</td>
<td>2.76</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>(.12)</td>
</tr>
<tr>
<td>Animals</td>
<td>M</td>
<td>2.91</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>(.40)</td>
</tr>
</tbody>
</table>
older subjects were using a more restricted definition of "animal" than the young subjects, i.e. they were searching a smaller domain.

To explore this possibility, we arbitrarily divided the animal category into subcategories (i.e. mammals, birds, reptiles and amphibians, fish, and insects) and then counted the number of items produced in each subcategory by the young and old subjects used in Experiment 1. These data are displayed in Table 9.

All the subjects reported examples from the mammal subcategory, and old and young subjects produced about as many examples from it. However, while all the young subjects (except Y-14, who reported no insects) reported examples from each of the other four subcategories, three of the old subjects produced only mammals, one produced no fish or insects, and one produced no insects. These data strongly suggest that old subjects, as a group, produced fewer clusters of animals overall because they tended to restrict their search to mammals.

Repetitions

All calculations in this experiment were based on using the total words produced by subjects, ignoring that some of the items were repetitions of earlier items. Gruenewald andLockhead discovered that, for their young sample, the percent of a subject's items that were repetitions was small, ranging from 0% to 7.7%, with a median of 0%.

However, in the present experiment it was discovered that the elderly subjects repeated items significantly more frequently than did the
Table 9

Number of Animals Recalled by Subcategory for Young and Old Subjects Used in Experiment 1

<table>
<thead>
<tr>
<th>Subject</th>
<th>Mammals</th>
<th>Birds</th>
<th>Reptiles/Amphibians</th>
<th>Fish</th>
<th>Insects</th>
<th>Other</th>
<th>Total Animals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y1</td>
<td>67</td>
<td>21</td>
<td>9</td>
<td>17</td>
<td>13</td>
<td>3</td>
<td>130</td>
</tr>
<tr>
<td>Y6</td>
<td>45</td>
<td>12</td>
<td>10</td>
<td>10</td>
<td>4</td>
<td>2</td>
<td>83</td>
</tr>
<tr>
<td>Y7</td>
<td>69</td>
<td>12</td>
<td>11</td>
<td>18</td>
<td>10</td>
<td>1</td>
<td>121</td>
</tr>
<tr>
<td>Y11</td>
<td>37</td>
<td>97</td>
<td>26</td>
<td>48</td>
<td>12</td>
<td>10</td>
<td>231</td>
</tr>
<tr>
<td>Y14</td>
<td>48</td>
<td>13</td>
<td>8</td>
<td>7</td>
<td>0</td>
<td>2</td>
<td>78</td>
</tr>
<tr>
<td>Y16</td>
<td>37</td>
<td>12</td>
<td>4</td>
<td>12</td>
<td>7</td>
<td>3</td>
<td>75</td>
</tr>
<tr>
<td>Mean</td>
<td>50.5</td>
<td>27.8</td>
<td>11.3</td>
<td>18.7</td>
<td>7.7</td>
<td>3.5</td>
<td>119.7</td>
</tr>
<tr>
<td>O3</td>
<td>48</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>48</td>
</tr>
<tr>
<td>O10</td>
<td>39</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>39</td>
</tr>
<tr>
<td>O13</td>
<td>60</td>
<td>31</td>
<td>9</td>
<td>23</td>
<td>16</td>
<td>0</td>
<td>139</td>
</tr>
<tr>
<td>O21</td>
<td>62</td>
<td>16</td>
<td>12</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>99</td>
</tr>
<tr>
<td>O22</td>
<td>71</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>71</td>
</tr>
<tr>
<td>O29</td>
<td>47</td>
<td>9</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>59</td>
</tr>
<tr>
<td>Mean</td>
<td>54.5</td>
<td>9.3</td>
<td>4.0</td>
<td>5.3</td>
<td>2.7</td>
<td>0</td>
<td>75.8</td>
</tr>
</tbody>
</table>

*Other* includes microscopic organisms and words which could not be understood from the tape recording.
young. Table 10 shows the average percent repetitions by age and category. An analysis of variance confirmed a significant effect for age, $F(1, 20) = 7.97, p < .02$. In the discussion section we will suggest that this age-related difference in percent repetitions is due to an episodic recall component in the semantic free recall task.

Table 10

Mean Percent Repetitions by Age and Category

<table>
<thead>
<tr>
<th>Category</th>
<th>Young Mean</th>
<th>Old Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Food</td>
<td>.87</td>
<td>3.93</td>
</tr>
<tr>
<td>Animals</td>
<td>1.85</td>
<td>6.58</td>
</tr>
</tbody>
</table>

Experiment 2

Manipulation Check

After the two experimental trials, each old subject was asked whether he had used the cues and whether they had been helpful. All of the subjects reported looking at the cues at some time during the experiment, but only 50% reported that they actually had found the cues helpful in recalling more items. This 50% was equally distributed over all the experimental conditions. That is, three subjects out of six in each condition reported that the cues were helpful, at least to some extent.
Although half the subjects reported that the cues had not been helpful during the experiment, all subjects were able to use the semantic differential cues to give exemplars from their cued category. After the experimental trials, each subject was asked to give three examples of foods or animals for each of the six semantic differential dimension cues. Subjects reported an average of 16.2 examples out of a possible 18 with a range from 15-18. For the animal category, an average of 4.7 of the examples were new items (i.e. they had not been reported during the experimental trial), while for foods an average of 5.0 of the examples were new.

BIIT's

To determine the effect of cuing on BIIT's, a 2 (category cued) x 2 (trial cued) x 2 (cued vs. non-cued) x 3 (time interval) factor analysis of variance was performed using mean BIIT in each 5-minute interval for each subject as the dependent measure. A significant main effect for time interval was found, $F(2, 40) = 9.36, p < .0005$. The main effect for cuing approached significance, $F(1, 20) = 3.51, p < .075$. This main effect was qualified by significant interactions of cuing by category cued, $F(1, 20) = 12.26, p < .002$, and of cuing by category cued by interval, $F(2, 40) = 4.57, p < .016$. Thus, the analysis indicated that the effect of cuing on BIIT depended on which category the subject received in the cued condition and on how long he had been performing the task.
The cell means for the three-way interaction are displayed in Table 11. For foods, the BIIT's are nearly equal for corresponding intervals, regardless of cuing. Also, BIIT's for cued and non-cued animals are similar in Interval I. However, in Interval II, and to a greater extent in Interval III, BIIT's for animals are shorter when subjects were cued than when they were not. A Newman-Kuels test for post hoc comparisons of means indicated that only the BIIT values for animals in Interval III exceeded the critical mean difference (39.90) for significance at the .05 level.

The mean BIIT's for the young subjects used in Experiment 1 are also listed in Table 11. The BIIT's for non-cued older subjects on the animals category are larger in Intervals II and III than the corresponding values for the young subjects. However, under cued conditions with the animal category, the BIIT values for older subjects begin to approximate the non-cued values for younger subjects.

**WIIT's**

When a subject had no multiple item clusters in a given interval, his WIIT for that interval technically would be zero. Since most subjects who did have at least one multiple item cluster in each interval had WIIT's in the range of 2 to 4 seconds, group means calculated by including values of zero would be seriously distorted. Therefore, the seven subjects who had a WIIT of zero in an interval were excluded from the analysis of variance.
Table 11
Mean BIIT in Seconds by Category, Cuing Condition, and Interval for Elderly Subjects

<table>
<thead>
<tr>
<th>Category</th>
<th>Cuing Condition</th>
<th>Interval</th>
<th>I</th>
<th>II</th>
<th>III</th>
</tr>
</thead>
<tbody>
<tr>
<td>Food</td>
<td>Cued</td>
<td>M</td>
<td>6.12</td>
<td>10.20</td>
<td>14.76</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SD</td>
<td>(1.34)</td>
<td>(3.93)</td>
<td>(5.21)</td>
</tr>
<tr>
<td></td>
<td>Non-cued</td>
<td>M</td>
<td>6.90</td>
<td>13.20</td>
<td>16.20</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SD</td>
<td>(1.86)</td>
<td>(5.79)</td>
<td>(8.45)</td>
</tr>
<tr>
<td>Animals</td>
<td>Cued</td>
<td>M</td>
<td>10.08</td>
<td>20.58</td>
<td>32.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SD</td>
<td>(2.17)</td>
<td>(6.34)</td>
<td>(20.39)</td>
</tr>
<tr>
<td></td>
<td>Non-cued</td>
<td>M</td>
<td>8.88</td>
<td>33.96</td>
<td>75.30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SD</td>
<td>(1.69)</td>
<td>(39.15)</td>
<td>(93.76)</td>
</tr>
<tr>
<td>Young</td>
<td>(Non-cued)</td>
<td>M</td>
<td>9.07</td>
<td>16.42</td>
<td>22.28</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SD</td>
<td>(3.15)</td>
<td>(8.38)</td>
<td>(9.58)</td>
</tr>
</tbody>
</table>
for the WIIT's. This procedure seemed to be the most conservative test of the effect of cuing on WIIT's.

The only significant effect in the $2 \times 2 \times 2 \times 3$ analysis of variance was a main effect for time interval. The mean values for WIIT, collapsed over all conditions and displayed as a function of time interval, are found in Table 12. We observe a tendency for the WIIT to increase between Intervals I and II, and then to remain nearly constant. However, as was observed in Experiment 1, there is no large increase in the WIIT as a function of time on task. The values are generally small and nearly numerically constant. By comparing Table 11 and Table 12, we can see that, while the time between clusters tends to increase with time on task, the time between items within clusters tends to remain relatively small. Thus, with increasing time on task, subjects appear to have increasing difficulty locating clusters; but once a cluster is located, items within it are emitted at nearly the same rate, regardless of how long the subject has been doing the task.

Table 12

<table>
<thead>
<tr>
<th>Interval</th>
<th>I</th>
<th>II</th>
<th>III</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>1.90</td>
<td>2.35</td>
<td>2.45</td>
</tr>
<tr>
<td>SD</td>
<td>(.50)</td>
<td>(.82)</td>
<td>(.98)</td>
</tr>
</tbody>
</table>

Note. $n = 17$. 
Number of Clusters

The mean number of clusters produced as a function of category, cuing, and interval is displayed in Table 13. Examination of the total number of clusters produced as a function of interval reveals consistently more clusters for foods than for animals. There appears to be no general effect for cuing. An analysis of variance confirmed these observations. The category cued by cuing interaction was significant, $F(1, 20) = 50.94$, $p < .0001$. Because of the experimental design, this interaction reflects an overall difference between food and animal values. The main effect for interval was also significant, $F(2, 40) = 162.967$, $p < .0001$. However, the three-way interaction which would indicate that the food/animal difference depended on interval was not significant.

The same pattern of significance was found for the single item clusters. There was a general food/animal difference, $F(1, 20) = 29.19$, $p < .0001$, and a main effect for interval, $F(2, 40) = 65.61$, $p < .0001$. No effect of cuing on the number of single item clusters was found.

For multiple item clusters, the highest order significant interaction was of category cued by cuing by interval, $F(2, 40) = 7.48$, $p < .0017$. Examination of Table 13 reveals that, for foods, the number of clusters decreases regularly with time, but that for animals, most of the decrease takes place by the end of 5 minutes. No clear effect of cuing on this pattern of decrease can be observed.
Table 13

Mean Number of Clusters by Category, Cuing, and Interval for Elderly Subjects

<table>
<thead>
<tr>
<th></th>
<th>Total Clusters</th>
<th>Single Item Clusters</th>
<th>Multiple Item Clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I</td>
<td>II</td>
<td>III</td>
</tr>
<tr>
<td>Food Cued</td>
<td>M</td>
<td>44.3</td>
<td>28.9</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>(11.7)</td>
<td>(11.3)</td>
</tr>
<tr>
<td>Non-cued</td>
<td>M</td>
<td>40.8</td>
<td>23.8</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>(15.7)</td>
<td>(9.2)</td>
</tr>
<tr>
<td>Animals Cued</td>
<td>M</td>
<td>27.8</td>
<td>14.8</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>(8.0)</td>
<td>(5.5)</td>
</tr>
<tr>
<td>Non-cued</td>
<td>M</td>
<td>30.2</td>
<td>13.2</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>(6.1)</td>
<td>(7.1)</td>
</tr>
</tbody>
</table>
Cluster Size

The average multiple item cluster size was calculated for each subject in each condition by dividing the total words in clusters of size 2 or greater by the total number of clusters of size 2 or greater, both as indicated by the algorithm. The standard analysis of variance was then used to determine whether there was any influence of condition on the average cluster size.

There were seven subjects with no multiple item clusters in some intervals. These subjects were dropped from the analysis because they produced no data from which to calculate an individual mean multiple item cluster size.

A significant effect found was for the interaction of category cued by cuing, $F(1, 13) = 12.09, p < .004$. Because of the experimental design, this interaction reflects an overall difference between foods and animals. In general, the cluster size for the foods was larger than for animals (see Table 14). The interaction of trial cued by time interval was also significant, $F(2, 26) = 4.62, p < .019$. No attempt was made to interpret this latter finding since the individual cell means involved in the interaction contained both food and animal values, some of which were based on n's of 3 and 4.

Although the mean multiple item cluster size was larger for foods than for animals in Experiment 2, the lack of a main effect for interval suggests that the size of multiple item clusters does not vary with time on task.
Table 14

Mean Size of Multiple Item Clusters by Category and Cuing Condition for Elderly Subjects

<table>
<thead>
<tr>
<th>Category</th>
<th>Cued</th>
<th>Non-cued</th>
</tr>
</thead>
<tbody>
<tr>
<td>Foods</td>
<td>2.79</td>
<td>2.79</td>
</tr>
<tr>
<td></td>
<td>(.40)</td>
<td>(.38)</td>
</tr>
<tr>
<td>Animals</td>
<td>2.41</td>
<td>2.49</td>
</tr>
<tr>
<td></td>
<td>(.52)</td>
<td>(.47)</td>
</tr>
</tbody>
</table>
DISCUSSION

The work reported in this dissertation evolved from a curiosity about why people experienced temporary inability to recall information which they "knew," i.e. information which they were later able to recall. I had observed this phenomenon in myself and in brain-damaged patients with whom I worked clinically, and had read that it was a common memory complaint among older individuals (Lowenthal & Berkman, 1967).

In order to understand why this phenomenon occurred, and to find clinical techniques to reduce its frequency in individuals where it represented a major problem, a reliable way of measuring it was a necessary first step. Without reliable measures of the frequency and duration of the phenomenon there was no quantitative basis to support speculations on its mechanism, and no standard against which to measure the effectiveness of intervention techniques.

The Bousfield (1944) semantic free recall task, and the Gruenewald and Lockhead (Note 1) paradigm for monitoring the time course of recall in that task, provided an experimental paradigm and a measurement system to begin a quantitative description of the temporal features of difficulties in retrieval of well-learned information from semantic memory.
Temporally, free recall in the Bousfield task, and in the similar continuous association task used by Levin and Breznitz (1977), proceeded in a series of rapid bursts of items separated by longer periods containing single or no items. Data from Dean (1971), Levin and Breznitz (1977), and Gruenewald and Lockhead (Note 1) supported the view that these temporal bursts were composed of related items. That is, extended semantic free recall could be viewed as a process of accessing clusters of related items of information, punctuated by periods of temporary blocks. Thus, the Bousfield task presented the opportunity to monitor a series of temporary blocks in recall from semantic memory for an individual subject.

Gruenewald and Lockhead (Note 1) developed a technique to measure these blocks during an extended period of recall by capitalizing on the finding that times within clusters of related items (WIIT's) were significantly shorter than times between such clusters (BIIT's). Thus from a collection of the IIT's in a free recall task, the longer times representing temporary blocks could be extracted. In addition, the method produced measures of the size and number of clusters accessed. All these measures could be studied as a function of time on task.

An elderly sample was selected for application of the Gruenewald and Lockhead measurement paradigm because, although this group typically complains of increased difficulty in recalling semantic information, there are far fewer studies in the literature on age-related changes in
semantic recall than there are for episodic recall. In addition, elderly subjects appeared to be aided more than young subjects by cuing in recall from episodic memory, raising the possibility that cuing could also improve performance in recall from semantic memory for the elderly.

**Replication of Gruenewald and Lockhead's Slope-Difference Algorithm on Young Subjects**

The data presented above comparing Gruenewald and Lockhead's young subjects with the young subjects used in this experiment indicate that, when applied to subjects of similar age and educational level, the slope-difference algorithm yields replicable numerical results for the values of BIIT, WIIT, number of clusters, and cluster size for corresponding 5-minute intervals in the recall of foods and animals. The similarities in the temporal properties of the data in the two studies, together with data from Dean and Gruenewald and Lockhead which indicate that temporal bursts are composed of semantically similar items, suggest that the slope-difference algorithm provides a reliable method of measuring the temporary blocks which occur in the access of successive clusters of related items.

The data from both experiments indicate that, for young subjects, (a) the number of words emitted is related to the number of clusters accessed, (b) the time to retrieve a new cluster increases with time on task, and (c) average size of a cluster does not vary significantly over the 15-minute period. In addition, the data indicate that time between
items within clusters increases only slightly with time on task. The fact that BIIT's and WIIT's change at different rates as a function of time supports the view that the retrieval processes involved in extended semantic free recall are of two types—retrieval of new clusters, which becomes progressively more difficult with time; and output of items within a cluster, which increases only slightly in difficulty with time on task.

**Young/Old Comparisons**

Neither the young nor the old subjects used in this experiment can be considered representative of their age group in general. Any conclusions drawn concerning similarities or differences in performance as a function of age can only be generalized to groups of healthy, well-educated individuals.

The most striking aspect of the data comparing young and old subjects on the Bousfield task is the similarity in performance between the two groups. When searching foods, the easier of the two categories (i.e. more total items are recalled), young and old subjects have equivalent mean BIIT's and show a similar increase in the mean BIIT over the 15 minutes allowed for recall from the category.

Although WIIT's were slightly longer (19%) for old subjects than for young subjects for both categories, this difference may not reflect a slowing in memory search rate. Older subjects did emit words in clusters more slowly. However, they also read words from a list 19% more slowly than did young subjects. Kirsner (1972) suggested that reading a
word from a list eliminates the memory search time involved in retrieving the word from long-term memory. Thus, the age difference in WIIT's can be attributed most parsimoniously to slower vocalization rates for the older subjects.

Old and young subjects also have equivalent BIIT's, number of clusters, and average cluster size during the first 5 minutes of the animal task. Only after the first 5 minutes do we observe longer BIIT's for the old subjects.

Thus, we are drawn to the conclusion that a healthy, well-educated sample of older individuals can access successive clusters of information in semantic memory as rapidly as young people, at least when the domain from which examples are drawn is large. This conclusion is further supported by the lack of young/old differences in mean verbal fluency for either "S" words (young = 19.1, old = 19.7) or "M" words (young = 16.6, old = 17.2).

The exception to the above conclusion is found when the old group recalls animals. Under these conditions, after 5 minutes on the task some of the old subjects show large increases in BIIT. How can we explain this increase?

When the number of animals recalled is expressed as the number of mammals, birds, fish, reptiles/amphibians, and insects recalled, we find that fewer of these "natural" subcategories were used by older subjects. These data raise the possibility that, if old subjects would have
been explicitly told to include examples of those subcategories, their total recall would have increased. We cannot know whether the increases would have brought them to the level of the young subjects, but it appears likely that the young/old difference in words recalled would have been reduced. Further work with such a change in experimental instructions would clarify this question.

We do know that the old subjects' deficit on the animal category was related to their recalling fewer multiple item clusters later in the task, and that when a multiple item cluster was accessed it was, on the average, the same size for young and old subjects. Thus, if the older subjects had used a less restricted definition of "animal," we could expect to see more multiple clusters accessed with a resulting decrease or elimination of (a) the age difference in percent of single item clusters and (b) the age difference in mean BIIT for animals in Intervals II and III.

Thus, when we examine the questions raised in the Introduction about whether old subjects have more difficulty searching semantic memory than young subjects (as they appear to do for episodic memory), we find little support for general age differences. The findings of Drachman and Leavitt (1972) now can be extended to show no difference in recall amounts up to 5 minutes for both categories used, although an age deficit does appear for animals after 5 minutes.

We do observe the greater within-group variability with age found in the work of Botwinick and Storandt (1974) and Thomas, Fozard, and
Waugh (1977). Baltes, Cornelius, and Nesselroade (Note 2) suggest that such variability reflects the more varied learning experiences which become available as age increases. The typical individual who is 70 has had much greater opportunity for a more varied learning history than a typical individual of 20. To the extent that the contents of semantic memory and the mechanisms involved in searching it reflect learning history, such increased within-group variability in performance might be expected.

A major question remains unaddressed by the conclusion of essential equality of performance in retrieval rates from semantic memory between young and old subjects. How do we explain the increase in the number of repetitions as a function of age?

Recent work by Ebner (1978) suggests that this increased tendency to repeat is not a function of the task used in the present experiment, but is a general feature of memory changes with increasing age. She gave young and old subjects a continuous recognition task, requiring subjects to examine cards on which were printed 200 words. Of the 200, 40 items were repeated twice with lags between the repetitions of from 2 to 32 words. Both young and old subjects showed an increasing tendency to fail to recognize a word correctly as repeated as the lag between presentations increased. However, for all lag periods, old subjects incorrectly reported a repeated word as a new word about twice as often as did young subjects.
Subjects in the present experiment can be viewed as having engaged in a continuous recognition task during recall. The rather small number of repetitions, along with informal comments by many of the old subjects during debriefing that they had tried not to repeat themselves, suggests that subjects understood the task as trying to produce new items. At the extreme, no subject simply said the same word over and over again. If we assume that during the free recall task there is some tendency for a previously reported word to be reaccessed (as suggested in a model of retrieval advanced by Shiffrin, 1970), we find that there is an episodic component to this semantic search task. Not only must subjects access category examples from semantic memory, they must remember whether they have said the word aloud before. Essentially they perform a recognition task. Ebner's data suggest that older subjects have more difficulty correctly recognizing a word as repeated than do young subjects. (She also observed no age differences in false positives on the task. New words were correctly indicated as such at equal rates for the two age groups.)

Thus, both Ebner's data and the data from this experiment are consistent with the hypothesis that under difficult recognition conditions (which Ebner defines as conditions where the subject is not allowed a forced multiple choice of whether one of two words in a pair has been seen before) recognition performance declines with age.

In the Introduction we suggested that the age-related difference in
recall from episodic memory might also be observed in recall from semantic memory if (a) more demanding tasks were utilized which removed the cuing inherent in multiple choice tests and (b) if times to emit a series of clusters, as opposed to single clusters or items, were measured. Neither of these suggestions were supported by the data in this experiment. Experiment 1 does not support the existence of major age-related temporal differences in semantic recall, at least when the number of items available for access is large.

Differences in performance on episodic and semantic recall tasks in the elderly further support the theoretical utility of Tulving's episodic/semantic memory distinction. Aging does appear to reduce the ability to recall information from episodic memory, or at least that aspect of episodic memory measured by laboratory free recall tests. However, the present experiment does not provide evidence for an age-related difference in retrieval times from semantic memory, at least when the semantic recall task involves producing a limited number of items from a large domain.

**Cuing Effects on Retrieval from Semantic Memory**

Research cited in the Introduction supports the general conclusion that recall from episodic memory is aided by semantic cues to a greater extent for the elderly than for younger subjects, although cuing increases recall in both groups. We suggested that, since in both episodic and semantic recall output appears to proceed by groups of related items of
information, if cuing increases the number of clusters of information recalled in episodic tasks, it might increase the amount of information recalled in semantic memory tasks as well. The results of Experiment 2 provide limited support for this hypothesis.

Only half of the old subjects given the cuing instructions reported that they were aware of utilizing the cues to aid recall, although all subjects reported that they examined the cues at some point during the recall process.

Other experimenters also have observed that elderly subjects tend not to use memory techniques in laboratory experiments unless explicitly required to do so. Most of this work has involved using imagery mediators to learn lists of words or items of information. For example, Hellebusch (1976) found that old subjects could improve recall of lists of words using a peg word mediation strategy when requested to do so. However, the old subjects failed to use the technique spontaneously in a second memory trial when they were not specifically requested to use the mnemonic aid. Robertson-Tchabo, Hausman, and Arenberg (1976) also reported the successful use of the method of loci (Yates, 1966) by older subjects when tested in the laboratory, but failure to use the method spontaneously when the subjects were tested in their homes.

The studies by Hellebusch and Robertson-Tchabo et al., plus the actual results of the cuing manipulation check in this experiment, indicate that when older subjects are given the opportunity to use a mnemonic
aid (in this case, semantic cues), but are not required to do so, there is a tendency not to utilize the mnemonic aid. In a discussion of memory skill training in the elderly, Poon, Fozard, and Walsh-Sweeney (1979) have stressed the importance of practice in the use of mnemonic aids for elderly subjects. In the present experiment, no practice was given in the use of the semantic cues before the actual experimental trials. It is possible that such practice would have increased the frequency of utilization of the cues.

We also observed that when subjects were required to utilize the cues to give category examples following the experiment, previously unreported items were accessed about 30% of the time.

At least two interpretations of this fact are possible. It might reflect the elderly subjects' tendency not to use mnemonic aids when left on their own. That is, cues might have been more helpful had they been more utilized. Alternatively, the recall of new items during the manipulation check might simply reflect a lack of exhaustive recall from the category during the experimental trial. Roedinger and Thorpe (1978) and Erderlyi, Finkelstein, Herrel, Miller, and Thomas (1976) have documented the ability of young subjects to produce additional new items in recall from 40 to 60 item lists, even after periods of 20 to 30 minutes. When Gruenewald and Lockhead (Note 1) allowed subjects to recall foods and animals for 30 minutes, as opposed to the 15 minutes used in this experiment, they also observed continuing, although slowing, recall
throughout the period.

The present experiment does not allow us to decide between these two possible explanations of cuing new items during the manipulation check. However, we can conclude that the semantic differential labels can function successfully as cues for the recall of new category examples, even for subjects who reported that the cues were not helpful.

When the effectiveness of cuing in semantic free recall was examined through the dependent measures employed in this study (i.e. BIIT, WIIT, number of clusters, and average cluster size), only the BIIT measure for the animal category was affected by cuing. The effect was observed, as expected, late in the recall period. There was no effect of cuing on WIIT, number of clusters, or cluster size. We also observed a decrease in the standard deviation of the group mean BIIT during those intervals where cuing reduced BIIT.

Although mean BIIT did drop under cued conditions, the lack of change in the mean number of clusters indicates that the actual effect of cuing on recall was minimal. Since the typical subject was producing very few clusters late in the task, a small increase in the number of clusters would appear as a large reduction in mean BIIT for that subject. If the clusters were evenly distributed over the interval, the mean BIIT would drop from about 250 seconds to about 125 seconds if the number of clusters changed from three to five. Since most subjects had mean BIIT's between 18 to 30 seconds for animals in Interval III, the impact of a small
change in the number of clusters for a few subjects would be to reduce greatly both the mean BIIT and standard deviation for the group. In effect, the BIIT measure magnifies a numerically small change in the number of clusters being produced when the clusters are being emitted at a slow rate.

The results of Experiment 2 indicate that, for older subjects, the impact of semantic-differential cuing, as used in this study, is much less striking than the impact of cuing on recall from episodic memory, as reported in the Introduction. We had originally suggested that cuing might aid semantic recall in the same way it appears to aid episodic recall, by reducing the difficulty in accessing the series of clusters of related items of information which characterize free recall. This hypothesis might have been more strongly confirmed if more older subjects had utilized the cues.

Our original decision to use semantic differential cues was inspired in part by Craik's (1968) observation of a progressively greater age difference in words recalled from a list as the domain from which that list was constructed was made larger. He suggested that older subjects were more benefited than young subjects by techniques which guided or constrained their memory search. The metaphor implied in this interpretation is one of searching a hypothetical space. Craik's interpretation of his data would suggest that old people are more apt to get stuck or lost in the space and therefore encounter fewer of the correct responses during their search. Alternatively, old people could be thought of as
searching the space more slowly, or as having longer distances between items in their space than young people.

Gruenewald and Lockhead's analysis of the Bousfield task also was based on the metaphor of searching a "semantic space," a hypothetical space in which the members of a category, such as animals, are arranged on the basis of their similarity. Since the semantic differential dimensions could be thought of as defining the dimensions of a general semantic space, we suggested that using the dimension labels as cues might induce subjects to move to new parts of the space if they were stuck in a particular region, or alternatively to move them more rapidly to new regions if the subjects' rate of search tended to be slow.

However, in the terms of this spatial metaphor, the results of Experiment 1 indicate that old subjects move through semantic space as easily as young subjects, encountering clusters of related items at equivalent rates. Age differences appear to arise not in the searching of a semantic space, but in defining that space. For young subjects, the "animal space" is more apt to include birds, fish, etc., while for old subjects there is a tendency for it to include primarily mammals.

An alternative interpretation would be that old subjects search a region of semantic space at the same rate as young subjects, but that the distances between regions are greater for the old subjects. Such an interpretation would imply that, for old subjects, the distance between, for example, mammals and insects would be greater than for young
subjects, increasing the probability that these regions would not be accessed by old subjects. Using the metaphor of greater distances between regions of semantic space, we can reinterpret Hultsch's (1975) work on the effect of category cuing in categorized free recall as helping old subjects traverse these greater distances more readily. Once a new region is accessed, the present experiment indicates that there should be no age-related difference in the ability to access items from within it, although there might be a difficulty in recognizing whether an accessed item had been heard recently during the list presentation. Such an interpretation would be consistent with Ebner's (1978) data on age-related differences on a continuous recognition task, and would account for the fact that, while Hultch's cuing procedure increased the number of clusters accessed, it did not increase the number of items emitted in a cluster.

**Evaluation of the Measurement Paradigm**

Two practical difficulties appeared in the use of the Gruenewald and Lockhead paradigm for studying the effects of manipulations on extended recall from semantic memory. First, the paradigm, as presently designed, requires many hours to process the data for an individual subject. Data-processing time could be reduced by having subjects press a key which signals a measuring clock each time a word is spoken. Also, access to higher speed computers could shorten the time-consuming curve fitting procedure. Second, large individual differences in rate of recall
exist for both young and old subjects, although the effect is more marked for the older group. These large individual differences require large groups of subjects to reduce error variance and therefore increase the ability of the analysis to detect the impact of manipulations that may have small magnitude of effect.

As indicated by the reanalysis of the young/old data within a repeated measure design, the tendency of a subject's performance on one category to be correlated with his performance on a second category can be utilized to reduce the problem of between-group variation.

However, the use of a repeated measure raises an additional problem not anticipated in the design of the experiments in this study. While we anticipated differences in the amount recalled from the food and animal categories based on Gruenewald and Lockhead's data, we did not anticipate that the two categories would show differential effects of the age and cuing manipulations. Future work which uses two different categories in a repeated measures design must take into account the possibility that the different categories might respond differently to experimental manipulations. In the present study and in Gruenewald and Lockhead's original work, the most salient difference in performance on the two categories is that many more foods are produced than animals. Although we cannot know whether this apparent difference in domain size is a critical variable or not, controlling for domain size would be a logical first step in subsequent use of a repeated measures design.
General Summary

The results of the above experiments indicated that when healthy, well-educated young and old subjects were asked to produce as many examples as possible from a semantic category (foods), they accessed new clusters, and emitted items within clusters, at the same rate as young subjects. Similarly, during the first 5 minutes of producing items from the animal category, no age-related differences were observed. It was suggested that increases in BIIT later in the task for the animal category reflected the fact that old subjects were more apt to produce only mammals as examples of animals, while young subjects also produced birds, fish, reptiles/amphibians, and insects. Thus, the study did not provide support for the existence of age-related differences in search rates for semantic memory.

Age-related differences were observed in the percent of items that were repetitions of previously reported items. It was suggested that this increased tendency to repeat an item was similar to Ebner's (1978) observation of an age-related deficit on a continuous recognition task.

Semantic differential cues resulted in shorter BIIT's for older subjects only for animals in the last 5 minutes of the task. The manipulation check indicated that only half the older subjects reported finding the cues helpful, so the minimal impact of the cuing manipulation was not unexpected. It was suggested that more practice with the cues might have led to a more dramatic effect of the manipulation.
Finally, it was suggested that the metaphor of a semantic space could be used to understand both (a) the effectiveness of semantic cues in increasing recall from categorized lists for old subjects and (b) the tendency of old subjects to report only mammals when asked to name animals if the semantic distances between regions (representing different taxonomic categories) were greater for old subjects than for young subjects.
APPENDIX A

RECALL TASK INSTRUCTIONS

Subjects were given the following instructions:

Today I will be asking you to give me examples from general categories which I will supply for you. For example, I might ask you to tell me all the books you can think of. You might mention specific books such as Gone with the Wind or the Bible. You might also give me types of books such as cook books, encyclopedias, or fiction books.

Anything which you consider an example of the large category will be considered an acceptable answer. There are no right or wrong responses to this task. Some people become concerned over whether or not I would consider something an example of the category. I stress to you that anything that you consider an example will be counted as such.

As you go through the task, I would like you to keep two goals in mind. First, come up with as many examples as possible in the time allowed. Second, even if you feel stuck, continue to try to think of examples until I tell you to stop. Most people who do this task feel, at some point, that they have run out of examples or that they are stuck. However, if they continue to try to generate examples, more examples frequently will occur to them. In summary, let me stress again, try to produce as many examples from the category as you can, and continue to try to produce examples until I tell you to stop.

Before we begin the actual experiment, I will give you a practice category and ask you to spend 2 minutes with it. After the practice I will answer any questions which may have arisen as you did the task. Then I will give you two large categories, one at a time, and ask you to spend 15 minutes giving me examples from each of them. During those categories I will be out of the room.

Following the two large categories I will ask you to do four short tasks which should only take 5 or 10 minutes.

Questions?

I will be tape-recording all your responses so that I can analyze them later. What I would like on the tape is just the examples. Please
don't think out loud or describe the examples to me. When I am done analyzing the tapes, they will be erased.

Questions? OK, let's try the practice category. Remember, try to produce as many examples as possible, and keep trying to produce examples until I tell you to stop. OK, give me as many examples as you can of "furniture."

(If the subject vocalizes anything more than examples during the practice trial, I say) Remember, I would like you to say only the examples you think of. Do not comment on the examples or tell me how you thought of them. Any questions?

OK, the first category name is written on this piece of paper (paper is face down in front of the subject). When I step out of the room, turn the paper over, read the category or name aloud and then begin to give me examples. Remember, give as many examples as possible and continue to try to think of examples until I tell you to stop. Questions?
APPENDIX B

CUING INSTRUCTIONS

Before the cued condition for each subject, he was given the following instruction:

On this piece of paper (face down) are six phrases which may help you to think of more examples when you begin to have difficulty thinking of more yourself. Think of them as hints or aids to help you think of more examples. When you look at the phrases, you might think, "What does he mean by that?" My answer is that I have no special meaning in mind. The phrases will mean different things to different people. Let them mean whatever they mean to you so that they might help you think of more examples. So when you feel that you are slowing down or are stuck, look at the card and try to use the phrases to help you think of more examples.

Questions?
REFERENCE NOTES


REFERENCES


Laurence, M. W. A developmental look at the usefulness of list categorization as an aid to free recall. *Canadian Journal of Psychology*, 1967, 21, 153-165. (a)


BIOGRAPHY

Raymond William Horn was born on December 30, 1945, in Sioux Falls, South Dakota. He obtained a B.A. degree in chemistry from Grinnell College, Grinnell, Iowa, in 1968. In 1969 he received a M.S. in inorganic chemistry from Stanford University, where he was supported by a National Science Foundation Fellowship. Between 1969 and 1972 he taught chemistry at Wabash College, Crawfordsville, Indiana. In 1972 Mr. Horn began graduate studies in psychology in the Clinical Psychology Program at Duke University under the support of a U.S. Public Health Service clinical traineeship. He completed his clinical psychology internship at the Veterans Administration Hospital in Durham, North Carolina, in 1976. Since that time he has been employed as a staff psychologist at John Umstead Hospital in Butner, North Carolina. His dissertation work was supported by a predoctoral research fellowship from the Administration on Aging. Mr. Horn currently lives in Durham, North Carolina.