1-1-1971

Retrieval processes for organized semantic information.

John G. Seamon
University of Massachusetts Amherst

RETRIEVAL PROCESSES FOR ORGANIZED SEMANTIC INFORMATION

A Dissertation

by

John G. Seamon

Thesis submitted to the Graduate School of the

University of Massachusetts in partial fulfillment of the

requirements for the degree of

DOCTOR OF PHILOSOPHY

September 24, 1971

Major Subject: Psychology

RETRIEVAL PROCESSES FOR ORGANIZED SEMANTIC INFORMATION

A Dissertation

by

John G. Seamon

Approved as to style and content by:

(Chairman of Committee)

(Head of Department)

(Member)

(Member)

(Member)

September 24, 1971
(Date)

ACKNOWLEDGMENTS

I would like to express appreciation to my chairman, Dr. James I. Chumbley, for his time, advice, and encouragement during all phases of this project.

I would also like to thank Drs. Charles E. Clifton, Jerome L. Myers, and Alexander Pollatsek for their cooperation and support.

TABLE OF CONTENTS

LIST OF FIGURES v
LIST OF TABLES vi
LIST OF APPENDICES vii
ABSTRACT viii
INTRODUCTION 1
METHOD 14
RESULTS 17
    Error Rates 17
    Reaction Time 18
    Category Lags 22
DISCUSSION 25
REFERENCES 31
APPENDIX A 33
APPENDIX B 34
APPENDIX C 35
APPENDIX D 36
APPENDIX E 37

LIST OF FIGURES

Figure Page

1. Mean RT with $\pm$ 1σ for each response type over the number of categories 19
2. Mean RT with $\pm$ 1σ for each response type over category size 21
3. Mean RT with $\pm$ 1σ for each lag condition over category size 24
4. Mean RT with $\pm$ 1σ for each lag condition over the number of categories 26

LIST OF TABLES

Table Page

1. Reaction time and category lag 23

LIST OF APPENDICES

Appendix

Page

A. Serial-Recall and Controlled Rehearsal Instructions 33

B. Memory Probe Instructions 34

C. Automatic-Recite Discrimination Instructions 35

D. General Instructions 36

E. Stimulus Materials 37

ABSTRACT

14 Ss learned serial lists of 9, 12, 15 and 18 items. The lists were composed of groups of semantically related words with the number and size of the semantic categories independently varied. Memory was probed by simultaneously presenting S with a category name and an item position cue. The results show that recall latency increased significantly with an increase in the number and size of the categories. A multi-stage retrieval model is indicated with category recall preceding item recall. In addition, recalls were produced by two different search routines: a high-speed automatic process, and a slower, problem-solving process.

INTRODUCTION

Information-processing approaches to human memory are concerned with how information is encoded, stored, and retrieved. The purpose of this paper is to focus on the latter concern and attempt to specify some of the cognitive processes involved in remembering semantic information. The methodology used to study cognitive processes is called Stage Theory and is an old idea in psychology. Earliest work in this area was done by Donders (Sternberg, 1969) who, in 1868, proposed that the time between a stimulus and a response is filled with a series of processing stages that occur in succession. Each stage is assumed to take some measurable, albeit small, amount of time with reaction time (RT) the sum of the component times for all stages. To measure the duration of a processing stage Donders used the subtraction method which consists of measuring the time to complete two tasks that differ only in the presence of a single stage. The RT difference between the two tasks is used as an estimate of the additional stage.

In an extension of Donders' original work, Sternberg (1969, 1971) points out that the assumption of *pure insertion*, upon which the subtraction method is based, is tenable only when proof is offered that other processing stages are not altered by the insertion of an additional stage. Two methods are presented to deal with this problem. The first method is called *selective influence*. Instead of inserting a new stage into the task, a quantitative variable is applied which should have an effect specific to a single hypothesized stage. If the temporal duration of the hypothesized stage is altered by differing amounts of the independent variable, the existence of the stage is confirmed. *Selective influence* does not provide an estimate of the total duration of the stage, but does, however, show whether there is such a stage and what its relationship may be to other processing stages. To obtain a stage duration estimate an additivity test for the pure insertion assumption is necessary. A quantitative variable is applied to a stage shared by two tasks which differ by a single, additional stage. If the effect of the additional stage is an increase in RT by a constant amount over all levels of the quantitative variable, it may be concluded that the additional stage did not interact with the stage to which the variable was applied. Tests of additivity between the additional stage and all other stages with which it may interact are necessary before a claim of pure insertion is appropriate. Using RT, Stage Theory thus provides a testable methodology for the study of cognitive processes.

Experimental Approaches. Two distinct experimental approaches to the study of semantic information retrieval have emerged in recent years. One approach is to make use of a person's pre-existing or available store of information. A question is asked and the time necessary to make a True-False or Yes-No decision is noted. The second approach requires recall or recognition of new learned material with the time to remember measured. The variables manipulated in these approaches tend to be the number or size of the semantic categories used. A stage model is implied by the assumption that memory must be searched before a correct response is made, and that finding the correct category must precede finding the correct category instance. The literature for each of these approaches will be considered in turn.

Available Information Approach. One of the first studies to make use of a person's available store of information is Landauer and Freedman (1968). Ss were presented with a category name. followed by a word which Ss had to classify as belonging or not belonging to the category. Time to make a classification was measured with the relative size of the semantic category either small or large. Landauer and Freedman assumed that before making a correct classification, Ss would have to search the category named to see if the word to be classified was an appropriate category instance. An effect of category size on RT would suggest the nature of the category search. If the target work is compared serially with the words of a category in memory, it is expected that RTs will be a direct, linear function of category size. A finding of no differeance in the slopes of the RT functions for positive and negative responses over all category sizes would imply that the target word was compared with all category instances before a decision was made. The serial search would therefore be considered exhaustive. A self-terminating process is indicated if the slope of the RT function fo positive responses is half that of negatives, as, on the average, only half of the category will have to be searched before a positive match is made, while the entire category must be searched for negative instances.

A different retrieval model is suggested if RT is an increasing, but non-linear function of category size. If the RT slope is negatively accelerating, a parallel search is indicated in which all category instances are compared simultaneously with the target word. The comparison time for each category instance is viewed as a distribution of times with the probability of at least one instance having a long comparison time greater for large categories than small. Consequently, small categories will yield faster RTs than large categories, but RT will not be a linear function of category size.

A finding of no difference in RT over category size would imply either a parallel search with a constant and equivalent comparison time for all instances, or a direct access process. If each category instance had a unique address, it may be possible to simply check that address directly and not have to search through irrelevant instances. Negative instances would not have an address within the category probed.

To vary category size, Landauer and Freedman used nested categories e.g., ANIMALS and DOGS, where the larger category, by definition, included all of the smaller. Relative category size, rather than absolute size, is determined by this procedure. Results show that positive responses were faster than negative responses, and only RTs for negative responses increased with category size for all levels of nesting. There is, however, much variability in the category size comparisons at each nesting level. For positive responses RTs for the category DOG were 29 msec faster than those for ANIMAL, while RTs for NOUN were 49 msec slower than those for WORD. It may be that different strategies were employed by Ss that in some cases enabled them to bypass the category search. Ss, for example, might only have to determine if the target word had any meaning to classify it as belonging to the category WORD. The finding of no significant difference in RT over category size for positive responses might thus be attributable to inappropriate averaging of data. If the WORD - NOUN comparison is excluded, positive response RTs are, on the average, 20 msec faster for small categories than for large categories, while a mean difference of 100 msec in the same direction is observed for negative response RTs. These data are consistent with a serial self-terminating process or a parallel process with variable comparison times.

Collins and Quillian (1970) performed a partial replication of Landauer and Freedman. Category size was varied by using nested categories and categories with different word count frequencies in Roget's Thesaurus. They found no consistent relationship between category size and RT for positive or negative responses. These results may, in part, be due to discrepancies between E-defined and S-defined category relationships. DOGS and BIRDS were both considered as nested categories of ANIMALS, with BIRDS larger than DOGS because of a larger number of instances listed in the Thesaurus. RTs for DOGS were faster than

ANIMALS, but those for BIRDS were longer than ANIMALS. If Ss had not considered BIRDS as a subset of ANIMALS, but rather defined ANIMALS as four-legged creatures, then extra time would be necessary to resolve the discrepancy before making a correct response. A finding of no consistent effect of category size on RT would not be surprising.

The most thorough application of this approach to the study of semantic information retrieval is found in Meyer (1970). He presented Ss with sentences of the kind "All S are P" and "Some S are P" and varied the set relationship and the size of the S and P categories independently. Ss were required to decide if the sentences were true or false as quickly as possible. The effect of category size on RT was found to be a function of the S and P relationship. In the conditions most comparable to Landauer and Freedman and Collins and Quillian, it was found that when the S category was a subset of the P category, decreasing S size or increasing P size produced an increase in RT,while, when S and P were disjoint, only an increase in P size produced an increase in RT. Meyer suggests a two-stage model for making correct decisions. In Stage 1 Ss try to decide if the S and P categories are related. If no relationship is apparent, a "False" response is made. For "Some S are P" sentences only this first stage is necessary. The finding that RT was equally long for "Some S are P" and "All S are P" sentences when the set relationship was disjoint, but faster for "Some S are P" sentences when the categories intersected suggests a serial self-terminating comparison process of the category attributes. Moreover, RTs for "Some S are P" sentences varied inversely with the amount of overlap in the set relationship between the S and P categories. These findings are consistent with a model which assumes that Ss serially compare the category name of S with the category names that intersect P and stop when a match occurs.

Sentences "All S are P" in which the categories are related require a second stage before a decision can be made. Stage 2 data suggests that Ss may make comparisons of the S and P category attributes. Meyer indicates that Stage 2 is a self-terminating process, either serial or parallel with variable comparison times, as the duration of this stage, which was obtained by subtracting "Some S are P" times from "All S are P" times, decreased with a decrease in the size of the P category. This decrease is attributed to a reduction in the number of comparisons that must be made before a decision is reached. Meyer's work thus suggests the stages involved in this task and specifies, to some extent, the nature of the operations in each stage.

Freedman and Loftus (1971) used the available information approach in a recall task. Ss were shown a noun category followed either 0.5 or 5.0 sec later by a letter or an adjective. S had to fill-in an appropriate word as fast as possible e.g., "ANIMAL (time delay) Z" with ZEBRA a correct response, or "ANIMAL (time delay) SMALL" with CAT a correct response. Relative category size was determined by having a different group of S list as many category instances as t hey could in a specified time period to each category noun. The categories were then rank-ordered in terms of the number of words produced. Freedman and Loftus hypothesized that, if retrieval was serial, RT would increase with category size and decrease with the amount of time between category noun and cue since, presumably, a serial search could begin at the category noun is given.

A low rank-order correlation between category size and RT was found indicating no effect of size on RT. In addition, the amount of time separating category noun and cue had no effect. Freedman and Loftus argue that these data indicate that noun categories can be accessed directly and that serial scanning for long-term memory is inappropriate. These conclusions, however, are questionable. In the previously cited studies Ss were presented with a category name and a possible instance and required to make a classification response. The finding that RT is longer for large categories than small categories is used to indicate that a memory search of the category was performed to find a match. The Freedman and Loftus task is fundamentally different. Ss were given a category subset as a cue and asked to recall any appropriate instance. A search of the category instances was unnecessary as Ss could output the first instance encountered. These data cannot, therefore, be used as evidence against a serial retrieval model. Neither can the finding of no significant difference between the time periods separating the presentation of the category name and the letter or adjective cue. The assumption that S might serially scan memory on the basis of a category name appears to be an unreasonable strategy to expects Ss to engage in e.g., NOUN, WORD, ACTIVITY, etc., and therefore might not provide a realistic test of the hypothesis. Thus the conclusion that noun categories can be accessed directly seems unwarranted. Rather, these data suggest that noun category retrieval, by whatever process, is not affected by category size. Retrieval of specific information from within a category, however, does appear to be a function of category size (Landaurer & Freedman, 1968; Meyer, 1970).

New Material Approach. In contrast with the preceding approach which does not involve new learning, the second approach to be considered deals exclusively with new learning. Most of the studies to be reported in this section are descendants of Sternberg's (1966) study of short-term recognition memory. In that study Sternberg presented Ss with a short string of digits, followed by a probe stimulus a few seconds later. Ss were required to indicate whether the probe was a member of the preceding string of digits by making a Yes or No response. The makeup of the digit strings was varied on each trial to insure that only short-term memory (STM) was tested. When the size of the digit sequence was varied from one to six, RT was found to be a linear function of set size with equivalent slopes for the positive and negative response functions. These data indicate that item recognition in STM is a serial exhaustive process.

Clifton and Gutschera (1970) modified the basic task by using two-digit numbers as stimuli instead of single digits. The data show that the two-digit probe was not compared with each two-digit list member in a serial exhaustive fashion. Rather, it was found that a two-stage process was necessary. Stage 1 involves a serial comparison of the tens digit of the probe to the tens digits of the list items. If no match is found, the search is terminated and a "No" response is made. If, however, a tens digit match is found, Stage 2 is initiated in which the ones digit of the probe is serially compared with all the ones digits that are associated with the matched tens digit. Ss were apparently able to make use of the structure built into the list to reduce the amount of material to be searched to make an item recognition response.

A similar conclusion was reached by Naus, Glucksberg, and Ornstein (1971). Using the STM recognition task, they presented S with word lists which varied in length from one to eight items. The word lists were further varied into single and double category sets. The double category sets were half animals' and half girls' names, with the categories presented in blocked order. RT over list length for the single category lists indicated a serial exhaustive search. Double category lists, however, showed a 25% reduction in the slope of the RT function suggesting a serial self-terminating search between categories and a serial exhaustive search within categories. A 50% slope reduction would be necessary to infer that search was specific to the category probed. This finding argues against a direct access model of category retrieval. Naus et al. replotted the data in terms of the effective, rather than the presented, set size and found that the slopes for the single and double category lists were equivalent. This indicates that categorization does not affect the rate of search, but instead permits fewer items to be searched.

The above studies show that retrieval in STM is a serial process. Sternberg (1969) reports an experiment in which the generality of the STM scanning model is tested. At the start of the experiment Ss memorized a list of 1, 3 or 5 digits which constituted the positive set to be used over the course of the experiment. On each trial a different list of 7 letters was presented. A few seconds after the last letters *Ss* saw either a recall signal and attempted to recall the 7 letters just presented, or a test digit. *Ss* made a positive or negative recognition response to the test digit with RT recorded. The purpose of the varied letter set procedure was to occupy the *Ss*' STM to insure that the recognition responses required retrieval from long-term memory (LTM). A control condition utilizing the standard STM procedure for set sizes 1, 3, and 5 was used as a baseline. Data from both conditions were linear with respect to set size with no difference between slopes for positive and negative responses. The linear function for the control condition had an intercept of 336 msec with a slope of 57 msec. Retrieval from LTM was different as indicated by a linear function with an intercept of 467 msec and a slope of 105 msec. The reliability of these scanning rate estimates, however, are subject to question. The scanning rate of 57 msec per item obtained from the control condition is considerably higher than the 38-40 msec rate typically found. Sternberg attributes this discrepancy to slowness in one of the four *Ss* tested. A serial exhaustive process is proposed for both conditions with two additional stages, which precede the serial comparisons,hypothesized for LTM retrieval. In Stage 1 a search is made of LTM to locate the positive set. The intercept difference between the STM and LTM is felt to reflect the LTM search time. After the positive set is located it is serially transferred into STM during Stage 2, with the average transfer time per item obtained from the slope of the LTM function. Stage 3 follows with a serial exhaustive comparison of the positive set with the test probe. Information is not retrieved directly from LTM, but is retrieved into STM where it is subsequently serially scanned for the desired information. Of course, it may be argued that information is not returned to STM to be scanned, but is located in

LTM by a slower scanning process. This would require separate scanning mechanisms for STM and LTM, or the same system operating at different speeds within STM and LTM.

Juola, Fischler, Wood, and Atkinson (1971) also applied the Sternberg recognition task to LTM. Ss learned a serial list of 10, 18, or 26 unrelated words to a high level of mastery by the serial recall method. Results are not consistent with a direct access retrieval model as RTs increased significantly with list length for both positive and negative responses. Linear fits for RT over list length are not good, however, thereby making it difficult to discriminate a serial from a parallel process with variable comparison times.

Juola *et al.* further analyzed RT in terms of test lag which was defined as the number of intervening words since the last test of a given word. The results for the words of each list length over Lags 0 through 30 are quite irregular, but a tendency for RT to increase as a function is present. Juola *et al.* argue that a recognition response could be made either on the basis of a memory search or a familiarity decision with familiarity inversely related to lag length. If these processes are present, then the finding that RT increases over lag length may reflect averaging over separate processes. Recognition responses produced by familiarity decisions are assumed to be faster than those produced by a memory search and occur more frequently at shorter lags than at longer lags. As such, neither process is examined independently over lag length. RTs produced by memory search should not vary over lag length if the same retrieval processes are involved in each response.

Deficiencies of the Experimental Approaches. The above review makes readily apparent the major deficiencies in each of the existing approaches. For those studies that have used the Ss' available store of information there is a lack of control over the absolute size of the category size since size can only be specified in relation to other categories. Category size must be treated as a qualitative, rather than a quantitative variable. Furthermore, the methods used to determine relative size do not insure that all _S_s have the same size categories in their LTM. Specific testing of retrieval models with differing assumptions is difficult at best.

Problems of control for size or number of categories are typically not found in the second approach which requires new learning. Category size and number can be precisely stated and treated as quantitative variables. Experimental control is unquestionably better with this approach, but at a cost perhaps of the generality of the results. These experiments have used serially ordered information, but few, if any, theorists would suggest that this is representative of all of memory. The function of this approach seems to lie in its attempt to provide a baseline by specifying the structure of the information committed to memory to determine the effect on retrieval.

The Experiment. The present research is an attempt to extend the baseline of knowledge by studying the retrieval processes involved in LTM recall of new learned information. Long word lists comprised of semantic categories are used with the number and size of the categories independently varied to determine their effect on RT. A two-stage serial process is suggested by work cited previously for item recognition (Clifton & Gutschera, 1970; Naus _et al_., 1971) and by free recall studies with categorized lists that have shown category recall to precede item recall (Segal, 1969; Bower, Clark, Lesgold, & Winzenz, 1969). The effect of the number of categories variable should be restricted to the category recall stage, and the size of categories variable to the item recall stage. Given this simple two-stage process, a fairly large number of different retrieval models may be generated to make predictions about the effect on RT of the number of categories, the size of categories, the serial position of a category within a list, and the serial position of an item within a category.

Models, Definitions, and Predictions

I. Two-Stage Serial Search. A two-stage serial search involves a sequential examination of the semantic categories followed by a sequential examination of the items within the desired category. Information may be serially examined in either a fixed order with the same or a randomly chosen starting point, or a random order on each trial. The search is self-terminating if it is stopped when the desired category is examined, continued within that category, and stopped at the desired item. S always knows the desired item since he is given the category name and item position as a recall cue. In an exhaustive search the desired category is selected after all categories have been examined with the process repeated on the items within that category.

A two-stage serial search is indicated if RT is a linearly increasing function of the number of categories (Stage 1), and the size of the list categories (Stage 2). A self-terminating, fixed order search is suggested for Stages 1 and 2 if RTs increase linearly with the serial position of the category in the list and the position of the item in the category. No effect of serial position for either stage would imply a self-terminating search with a random starting point or random order, or an exhaustive process.

II. Two-Stage Parallel Search. A two-stage parallel process involves a simultaneous comparison of all list categories followed by a similar comparison of the contents of the desired category. Parallel searches can have constant or variable item comparison times and may be self-terminating or exhaustive. If the comparison times are constant (CCT) all categories can be accessed at the same time, and all items within the desired category at some constant, additional time. Variability in RTs is produced by motor movements and is independent of the number and size of the categories. If, however, comparison times are variable (VCT), search times are really a distribution of times with equal mean access times for all categories, and equivalent, but longer, average search times for random variable and variability should increase with the number and size of the categories. A self-terminating search terminates the category recall stage when the desired category is found and stops the item recall stage when the desired item within that category is located, while an exhaustive search is terminated only after all categories and all items within the selected category have been examined.

A two-stage parallel search may be indicated in a number of alternative ways. If RT remains constant with variation in the number and size of categories, and systematic effects of category position and item position within a category are not observed, a parallel model with CCT is suggested for Stages 1 and 2. If increments in category size produce a negatively accelerating effect on RT, while all other variables have no effect, a parallel search with CCT is indicated for Stage 1 and a parallel search with VCT for Stage 2. If a negatively accelerating effect is produced by increments in the number of categories, with all else constant, the reverse of the preceding model would be implied. Finally, a negatively accelerating effect on RT with increases in both the size and number of categories, and no systematic serial position effects of categories or items, suggests a parallel search with VCT for both stages.

III. Direct Access Process. If each list item in memory has a unique address it may be possible to bypass the memory search and access the information directly. This process is suggested if RT remains constant with changes in the independent variables. Since a two-stage parallel search with CCT makes the same predictions, these models are indistinguishable in the present context.

IV. Two-Stage Serial-Parallel Search. A serial-parallel search involves a sequential examination of the semantic categories followed by a simultaneous comparison of the contents of the desired category. Results suggesting a serial process for Stage 1 and a parallel process for Stage 2 would support this model.

V. Two-Stage Parallel-Serial Search. Reversing Model IV, a parallel-serial search requires a simultaneous comparison of all list categories in Stage 1, and a sequential examination of the items in the selected category in Stage 2.

Because of the varying assumptions all models are not completely discriminable in terms of the independent variables, however, serial models can be differentiated from parallel and direct access models, some parallel models can be differentiated from a direct access model, and simple models can be differentiated from complex models.

METHOD

Fourteen students from the University of Massachusetts served as paid Ss. Individual Ss participated for 90 mins on each of 4 consecutive days. During each daily session Ss learned and recalled two lists of words with a rest period separating each list. Payment was made at the completion of the experiment and was not conditional upon performance.

The experiment was conducted in a sound-proof environmental chamber. S and E sat in the chamber at separate desks and viewed experimental displays on individual TV monitors. Displays were controlled by a PDP-8/I computer located outside of the testing chamber. Instructions to the computer for changing displays and storing data were made by E on a response console within the chamber.

The first list learned in the opening session by each S contained three categories of three words each (3x3) viz., DUKE-PURDUE-CORNELL-LINDA-NANCY-JANE-ROME-PARIS-LONDON. It served as a practice list during which instructions were given in piecemeal fashion to avoid overburdening S. Prior to list learning, instructions for the serial-recall method and the controlled rehearsal procedure were presented (See Appendix A). Unlike typical serial learning experiments where one word at a time is presented for learning, this experiment presented all the words in a given category simultaneously. This methodological departure is based on the results of a series of experiments (Seamon, 1972) which indicate that RT is influenced by trace strength and that the influence can be equated by controlling rehearsal processes.

The words of a category were arranged horizontally on the TV monitors and displayed for a time period dependent upon category size. The display time was calculated by multiplying the number of words per category by 6 secs. Thus a three word category was shown for 18 secs, and a six word category was displayed for 36 secs. While the words of a category were being displayed, S was required to rehearse the group by reciting out-loud the entire group sequentially between auditory clicks that were provided by a solenoid switch and two Hunter timers. The clicks were temporally spaced to provide 0.6 sec to rehearse each word in the displayed category. The ratio of display time to rehearsal time permitted each word of a category to be rehearsed 10 times per trial. A three word category, for example, was shown for 18 secs and had an auditory click every 1.8 secs, while a six word category was shown for 36 secs and had a click every 3.6 secs.

After the last category in the list was presented and rehearsed, the word "READY" appeared on the screens. This word served as a signal for S to recall all the words in their correct serial order on a sheet of paper which had the appropriate number of blanks provided. Only the final serial arrangement of the words was important. S could fill-in the response sheet in any desired temporal order.

Following recall E removed the response sheet and initiated another trial of list presentation and serial-recall. This procedure was continued until the word list was recalled perfectly on two successive trials. Memory probe instructions were read after list mastery (See Appendix B). A memory probe consisted of the simultaneous presentation of a category name and an item position cue e.g., BODY _ X _ on the TV monitors. S was required to fill-in the word that occurred in the position signified by the X, in the category named, as quickly and as correctly as possible by making the response verbally into a voice key. Approximately 3 secs elapsed between memory probes with no feedback given. One presentation of the individual probes for all list words constituted a trial and was followed by a 30 sec rest period. A total of 20 trials was used, with the memory probes presented in a random order on each trial.

In previous research with this paradigm (Seamon, 1972) Ss frequently reported that some words were recalled in an "automatic" fashion without awareness of the retrieval process, while other words were recalled only after other members of the probed category were recalled. If the reports are accurate they would indicate that information may be retrieved by more than one type of process. To test this possibility Automatic-Recite discrimination instructions were read during a 3 min break between Trials 10 and 11 (See Apprendix C). A response of "Automatic" or "Recite" was required after the recalled response to each probe. S was instructed to say "Automatic" if his verbal response was the only word thought of when the probe was presented. If any other word or words came to mind he was told to say "Recite".

Upon completion of the first list, general instructions were read indicating that the preceding list was a practice list and that the procedures used during serial learning and Trials 11-20 of the memory probes would be used in the remainder of the experiment (See Appendix D). The 14 Ss were subsequentially run in a Latin Square design with seven orders of the experimental conditions and 2 Ss per order. The variables manipulated were the number and the size of the semantic categories in a list of words. Number of categories was varied from three to six while holding the number of categories to three. Seven experimental combinations of number and size of categories were obtained with one 9 item list (3x3), two 12 item lists (3x4, 4x3), two 15 item lists (3x5, 5x3), and two 18 item lists (3x6, 6x3). This procedure enabled the number and size of the categories to be varied while holding the total number of items in memory constant viz., Conditions 3x4 vs. 4x3, 3x5 vs. 5x3, and 3x6 vs. 6x3. The original 9 item practice list was not used in the data analysis.

Words for the experimental lists were obtained from the category norms of Battig and Montague (1969). Twenty-seven categories with six instances per category were used to construct the lists (See Appendix E). For each list within the seven orders of seven lists the categories and the particular category instances were selected randomly. Ss saw each of the 27 categories only one time and in a randomly determined arrangement. Only the 2 Ss from the same order of the Latin Square saw the same categories and category instances in the same list conditions. The possibility of these Ss learning identically ordered lists, however, is remote as each list was reordered when loaded into the computer by randomizing the order of the categories and the order of the words within a category.

RESULTS

The error rates for the seven experimental conditions had a range of 1.34% to 1.95% with a mean of 1.65%. Errors were low in each condition and did not vary systematically over the number and size of the categories. Errors, when made, were more likely to be Automatic responses (79.22%) than Recite responses (20.78%), with the latter occurring on the initial test trials. Subsequent analyses are based on correct responses only.

Reaction Time. - Median latencies were determined for Automatics and Recites for each S at each serial position with a mean, based on the medians, then obtained for each S for each response type and list condition. The individual S means were used in separate analyses for the number and size of categories variables. All analyses are based on items which occurred at the ends of each category only. Items from within a category are not included. Ss indicated that a problem in probe discrimination was present as category size was increased, e.g. _ x _ vs. _ _ x _ _ _ . This is supported by an examination of RT over serial position. For lists containing three word categories there was no systematic effect of category position within the list or item position within a category for either Automatics or Recites. As category size increased, however, a bowing effect within categories was observed for Automatics and Recites with end positions faster than middle positions. To the extent that probe discrimination was a problem, RT is increasingly inflated for categories of more than three words. Use of only end of category items eliminates this problem. All analyses,however, were performed using all serial positions as well as only end of category items. The results were identical in each instance.

Figure 1 shows RT for Automatics and Recites for the number of categories. RT for both functions increases in a step-like fashion with the major change occurring between list conditions 4x3 and 5x3. An analysis of variance for Response Type and Number of Categories yielded significant effects of Response Type (F 1, 13=220.65, $p<.001$) and Number of Categories (F 3, 39=5.73, $p<.01$) with the interaction not significant (F 3, 39=1.16, $p>.05$). RTs were longer for Recites than for Automatics and varied with the number of categories. A trend analysis revealed significant linear (F 1, 13=13.54, $p<.01$) and cubic

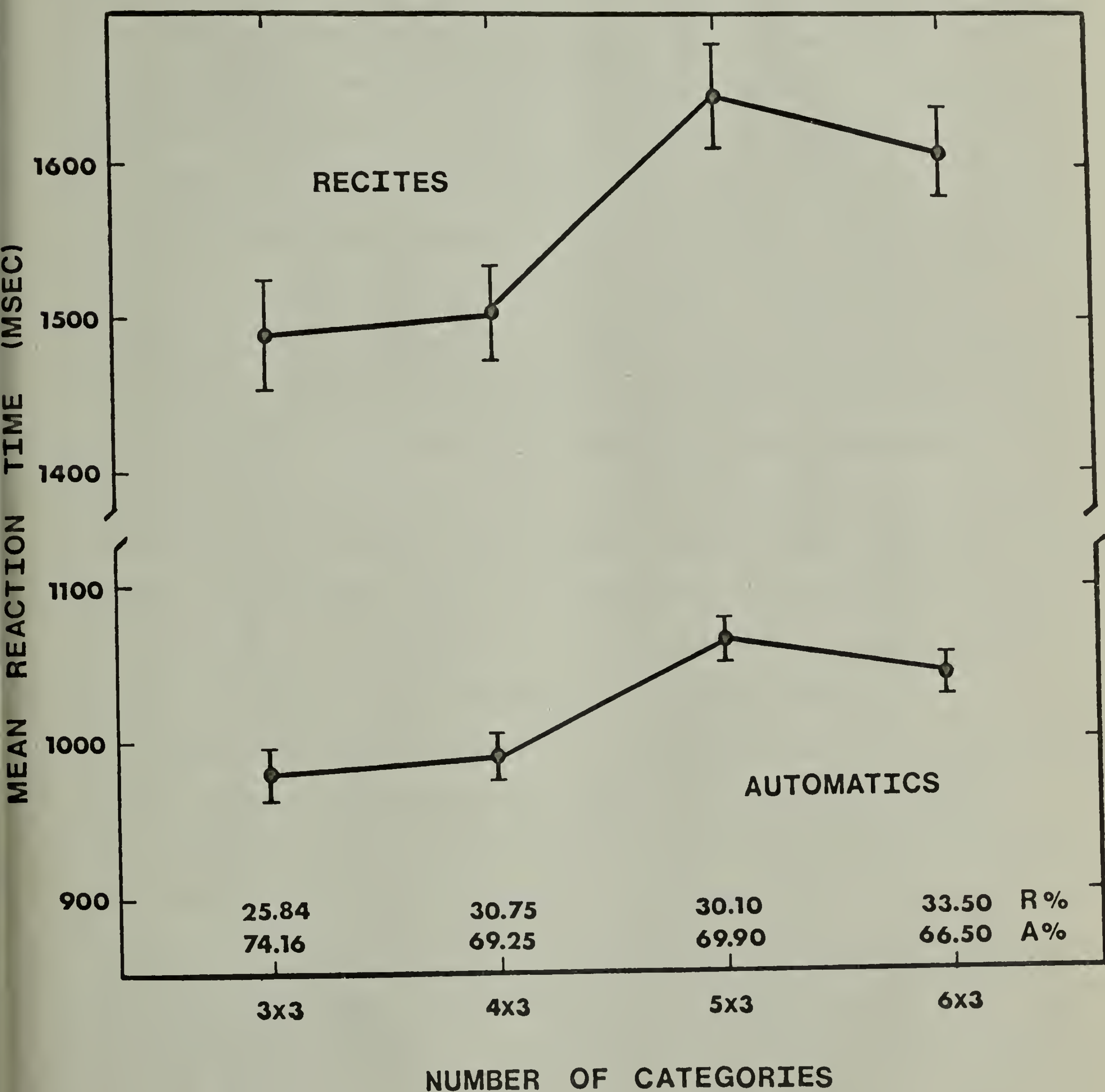

Fig. 1. Mean RT with ± 1σ for each response type over the number of categories.

(F 1, 13=13.94, p<.01) components for the Number of Categories, with the interaction of Response Type and Number of Categories not significant for either linear (F 1, 13=2.02, p>.05) or cubic (F 1, 13=2.28, p>.05) trend. An F value less than 1 was found for the quadratic component. Also shown in Figure 1 is the mean percent of Automatic and Recite responses. There is a tendency for Automatics to decrease and Recites to increase with the number of categories.

Figure 2 shows the mean RT for each category size and response type with the best-fitting linear functions. An analysis of variance for Response Type and Size of Categories produced significant effects of Response Type (F 1, 13=206.19, p<.001), Category Size (F 3, 39=15.04, p<.01) and their interaction (F 3, 39=6.02, p<.01). Recites were longer than Automatics and increased at a greater rate over category size than Automatics. A trend analysis found a significant linear trend for Category Size (F 1, 13=49.12, p<.001) and a significant linear interaction between Response Type and Category Size (F 1, 13=19.42, p<.001). Nonsignificant effects were obtained for the quadratic (F 1, 13=1.13, p>.05) and cubic (F<1.0) components. The mean percent of Automatics and Recites, as seen in Figure 2, is not systematically related to category size.

Separate Latin Square analyses of variance were performed on Automatic and Recite responses from the seven list conditions. The Automatic responses showed significant effects of List Condition (F 6, 72=8.78, p<.001) and Practice (F 6, 72=4.59, p<.001) with the interaction not significant (F<1.0). Analysis of the Recite responses revealed only significantly different List Conditions (F 6, 72=7.13, p<.001). Practice effects were not significant (F<1.0), nor was a Practice by List Condition Interaction significant (F<1.0). These findings suggest that Ss' retrieval processes were fairly constant over all list conditions. With practice Ss got faster at making Automatic responses, and the

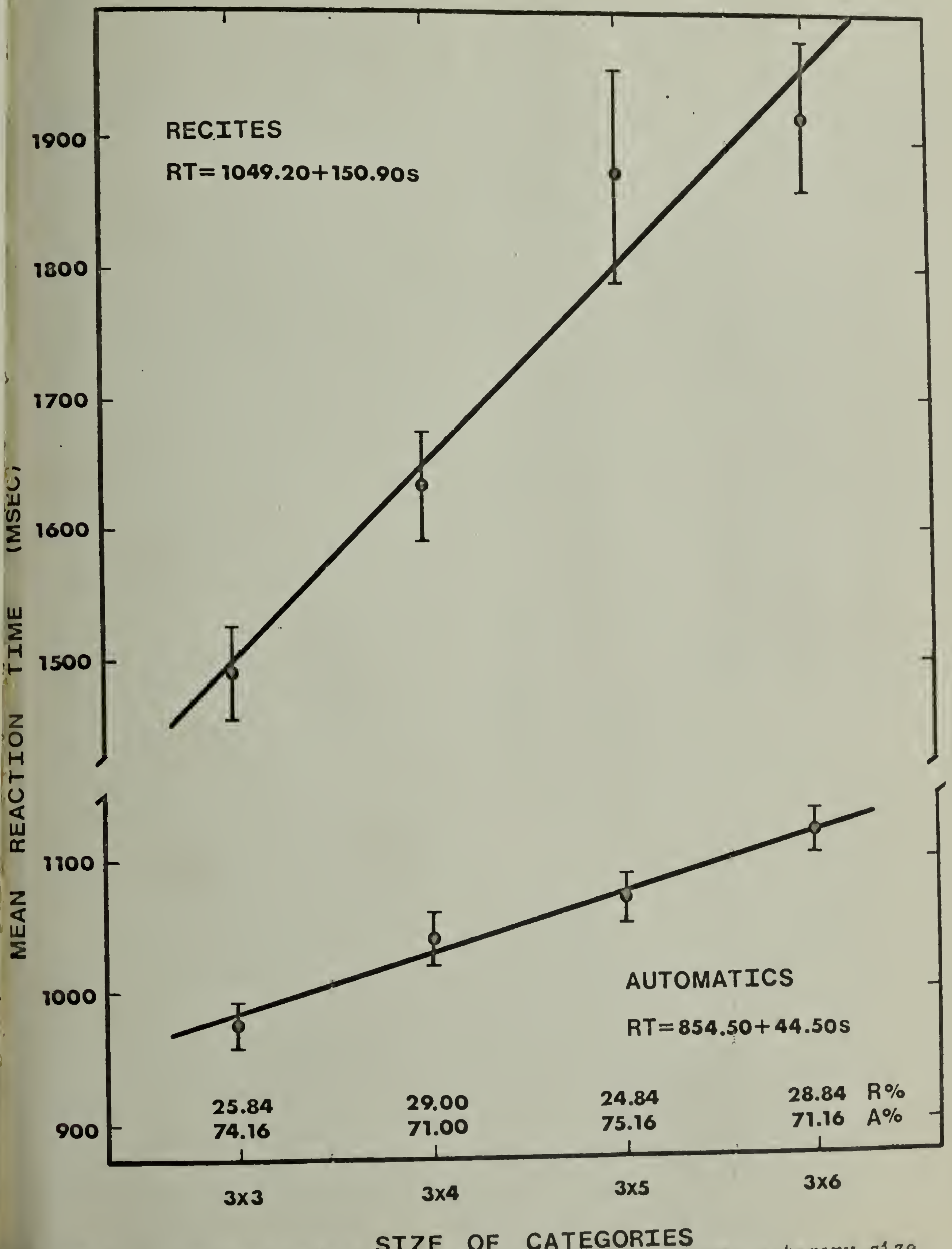

Fig. 2. Mean RT with ± 1 σ for each response type over category size.

facilitative effect of practice was equivalent for all lists. Recite responses, unlike Automatics, did not improve with practice. Data from post-experimental S reports indicate that Recite responses involved only instances from the probed category. Items from other categories were not examined.

Category Lags. - The use of a random order of memory probes permitted an analysis of RT as a function of category lag. Category lag is defined as the number of memory probes between presentations of the same category during a test trial. Mean RTs for lags of 0 to 10 were obtained by collapsing the individual RTs for all Ss for all end of category items. Table 1 presents the mean RT for each category lag over all lists. Of principle interest is the finding that lag length had little effect on RT with the exception of Automatics at Lag 0. These responses were clearly faster (99% confidence interval) than responses to lags of 1 or more which did not vary. Recite responses were longer than Automatics in each instance and were unaffected by lag length. This implies that the underlying retrieval processes for a Recite response were the same under all lag conditions, while the retrieval processes for an Automatic response were equivalent only for Lags 1-N.

Since RT did not vary as a function of lag from 1 to N for Automatics, these data were combined to yield more observations for an examination of RT by lag and list condition. Figure 3 presents these data as a function of category size. It is readily apparent that lag length interacted with category size. As category size increased, the difference between lag conditions decreased. The finding that the slopes for the Lag 0 and 1-N functions are larger than that reported previously for Automatics over all lags is not conflicting as the mean for each category size and lag length was based on the individual RTs from all Ss, while the earlier cited slope was based on the mean of the individual

TABLE 1

Reaction Time and Category Lag

	Automatics			Recites		
Lag	RT msec	99% CI	Frequency	RT msec	99% CI	Frequency
0	999	982-1016	1416	1750	1670-1830	472
1	1059	1040-1078	1054	1756	1677-1836	496
2	1055	1033-1077	797	1747	1646-1847	409
3	1055	1029-1080	619	1752	1659-1844	317
4	1056	1025-1087	417	1825	1686-1964	224
5	1050	1014-1086	205	1765	1561-1968	132
6	1088	1033-1142	134	1737	1592-1881	108
7	1062	1003-1121	93	1754	1521-1987	51
8	1037	964-1110	61	1743	1460-2026	25
9	1092	996-1189	36	1524	1319-1729	11
10	1074	938-1211	14	1943	1390-2495	8

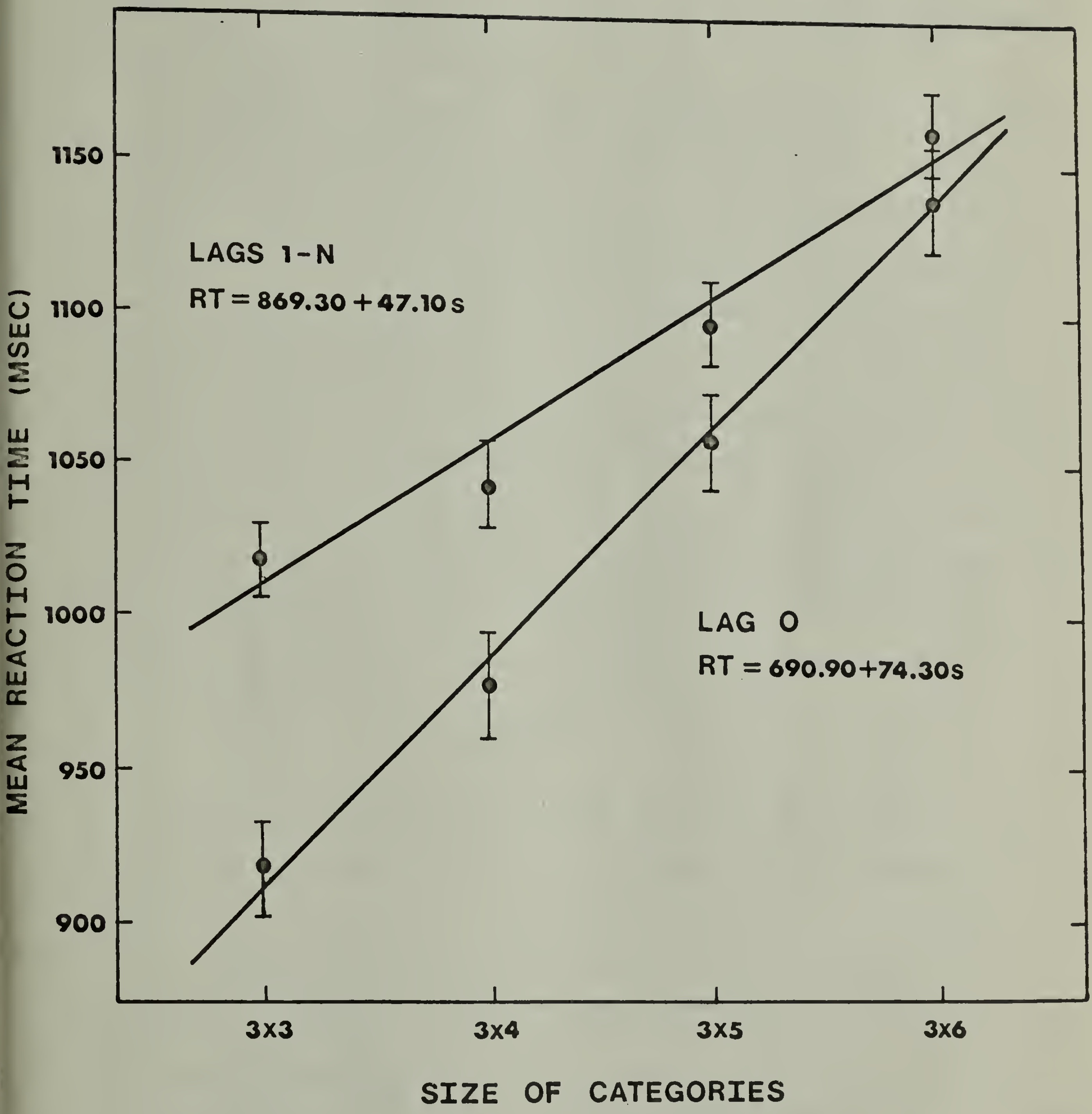

Fig. 3. Mean RT with 1 for each lag condition over category size.

S means for each category size. Use of this procedure permits the frequency of occurrence of each lag length to be observed. The former method, however, equates the influence of each S on the group statistic and is therefore more appropriate for parameter estimation.

The RT data for lags over the number of categories are shown in Figure 4. To determine the best-fitting equations, STEPIT (Chandler, 1965), a search algorithm, was used to obtain linear, quadratic, and cubic functions for RT over category number and lag. As shown in Figure 4, all functions yielded values within $\pm 1\sigma$ for Lag 0, but only the cubic function was consistently within this limit for Lags 1-N. The cubic functions, however, drop with an increase in the number of categories and are, therefore, difficult to interpret psychologically. The quadratic functions are likewise difficult to interpret as a negatively accelerating function was obtained for Lag 0 and a positively accelerating function for Lags 1-N. Finally, the linear functions, while showing RT to increase with the number of categories at both lag conditions, are also lacking as they provide poor fits of the data, especially at Lags 1-N.

DISCUSSION

This study has shown that information can be recalled by more than one type of retrieval process, and that the number and size of the semantic categories within a list committed to memory have significant and different effects on RT. Differential effects are indicated by a) a linear increase in RT with an increase in category size, and a smaller, non-monotonic increase in RT with an increase in category number, and b) a significant interaction of response type with category size, but not with category number. These data do not support a simple trace strength explanation which posits RT to be a function of the

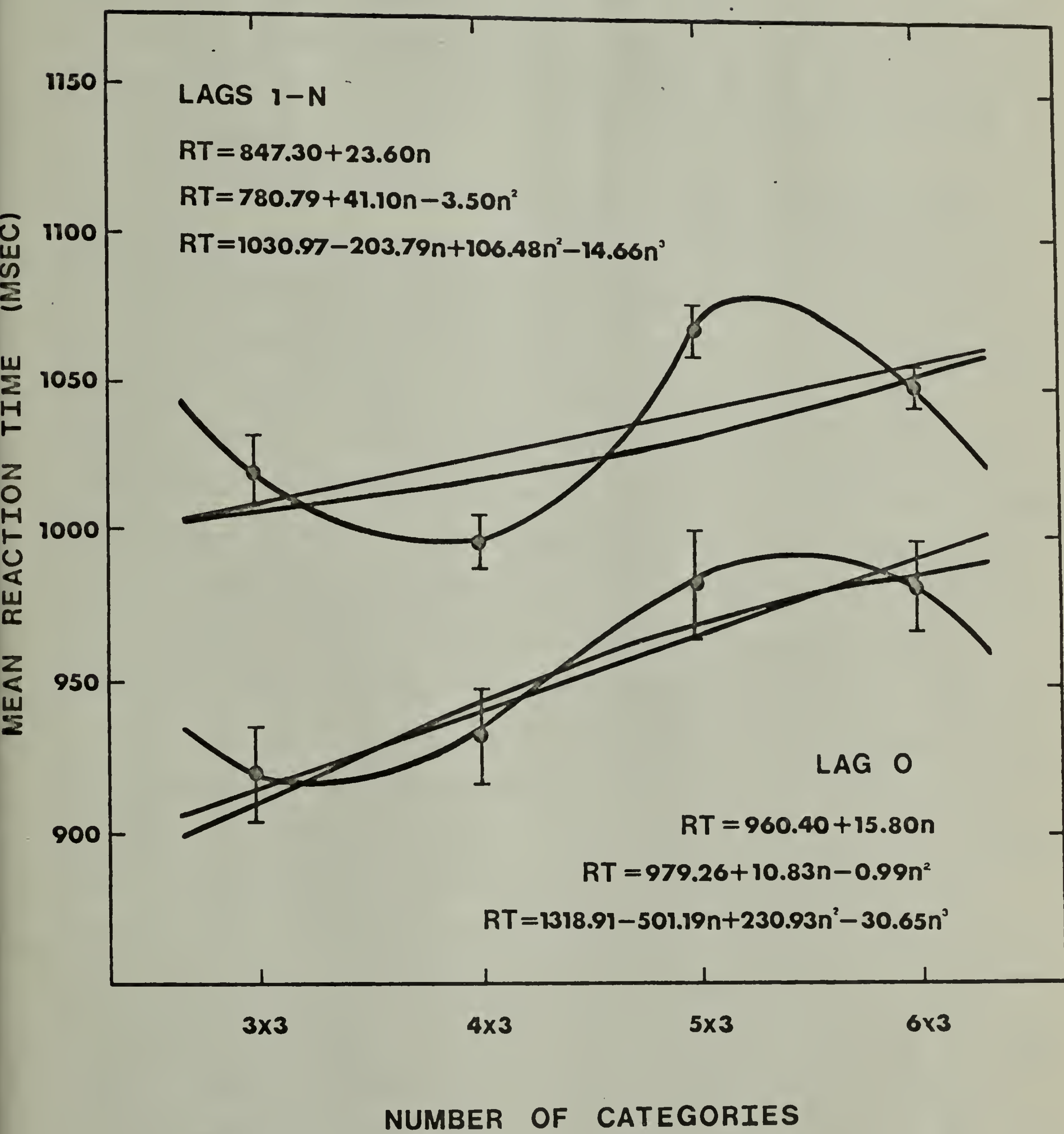

Fig. 4. Mean RT with 1 for each lag condition over the number of categories.

number of items in memory. Differences between Lists 3x4/4x3, 3x5/5x3, and 3x6/6x3 indicate that it is not the number of items which influences RT, but the structure or organization of the items within the list. This suggests that the independent variables exerted their influence on different processing stages during retrieval, and are consistent with the hypothesis that category recall precedes item recall.

One account of the data suggests that on a given trial S first reads the display to determine the category and item position being probed. The desired category is retrieved by either a serial or parallel process with variable comparison time. The non-monotonic increase in RT over the number of categories makes it difficult to differentiate these processes, but does permit rejection of any model which assumes that a category can be accessed immediately. After the category has been retrieved, the desired item is obtained by either an Automatic or Recite search routine. The linear increase in RT over category size for Automatics and Recites indicates a serial process for both routines. It is not known what determines which routine is used. It may be that probe discrimination or initial fixation on the display screen plays a role. In the Automatic routine S performs a high-speed serial scan of the items from the retrieved category at a rate of 11 to 22 words per second depending upon whether the process is self-terminating or exhaustive. The scan rates are obtained from the slope of the RT function for Automatics over category size (44.50 msec) which is essentially the same as that for Automatics over category size at Lags 1-N (47.10 msec). If the process is exhaustive, the scan rate is equal to the slope, while if it is self-terminating, the slope must be doubled, as on the average, only half of the items will have to be scanned before the correct response is found. The data do not permit differentiation of these assumptions since the finding that RT did not increase with serial position could reflect an exhaustive scan or a self-terminating scan with a random starting point or random order. After the desired item is obtained, it is emitted followed by a response of "Automatic".

In the Recite routine a much slower serial scan of the words is performed with a rate of 3 to 6 words per second again depending upon whether a self-terminating or exhaustive assumption is appropriate. Either would yield a rate within Landauer's (1962) estimate of the range of subvocal rehearsal of 3 to 7 items per second which suggests that this might underlie Recite responses in the present experiment. After an item has been obtained, it is checked to insure that the retrieved position matches the probe position. If a match is obtained, the word is emitted followed by the word "Recite".

Since the category recall stage is the same for Automatics and Recites, the model would not predict an interaction of response type with the number of categories. However, an interaction would be predicted for the item recall stage as Automatics and Recites have different processing rates. Recites are always longer than Automatics because of the presence of a retrieval check which, while consuming time, improves accuracy. Moreover, describing Recites as subvocal recitation suggests that these responses would not improve with practice if S was reciting as quickly as possible at all times.

After making a response it is not unreasonable to assume that S has information available from that probe. This information could be all or some of the items from the previously probed category in active memory. Research from running memory (Atkinson & Shiffrin, 1968) and free recall (Rundus & Atkinson, 1970; Glanzer, 1971) studies indicate that the capacity of active memory is typically 3 or 4 words. S might have retained 3 or 4 items from the previously probed category in active memory and eliminated the necessity of the category recall stage when one of those items was probed on the succeeding trial. This advantage would be restricted to Automatics as the retrieval check which is hypothesized to follow Recites could knock information out of active memory.

If S did employ a strategy of retaining items from the previously probed category this would predict RT for Lag 0 to be faster than Lags 1-N due to the possible elimination of the category recall stage. RT should not vary over Lags 1-N, and the effect should not be present for Recites as information is not available to bypass the category recall stage. The interaction of lag condition with category size reflects the limited capacity of active memory. As category size increases, the probability of the probe item being in active memory decreases, and RT for Lag 0 approaches that for Lags 1-N. Further, the use of this strategy might vary with the probability of a Lag 0 which decreased as the number of categories increased (32.13, 24.57, 21.17, and 17.23%). A flat RT function over the number of categories at Lag 0 might therefore not be expected.

The finding that the act of recall can involve more than one type of retrieval process seems to conflict with Sternberg (1967) who reports a serial self-terminating process with a scan rate of 4 items per second for a short-term recall task. Sternberg, however, did not differentiate Automatics from Recites and thus these processes may have been present, but undetected. Alternatively, it may be that the task of recalling the item which followed the probe induced S to adopt a strategy of subvocally reciting the serial items to find the correct response. Sternberg's scan rate is very similar to that observed in the present experiment for Recites. The specification of different available retrieval processes agrees with the view of retrieval from long-term storage as a multi-process activity (Reitmen, 1970; Feigenbaum, 1970) in which information may be retrieved by different processes at different times.

The present experiment and previously cited research have indicated that serial processes are present in both recall and recognition and for short-term and long-term storage. The argument frequently made against the generality of a serial retrieval model (Landauer & Freedman, 1968) is that it would be too time-consuming for large amounts of information. This would appear to apply, however, only if the information is unorganized. The present results and others (Clifton & Gutschera, 1971; Naus *et al.*, 1971) clearly show that *S* makes use of whatever organizational properties are present to reduce the amount of material to be searched. It might be more fruitful to examine the generality of this model, not in terms of the quantity of information, but the manner in which it is processed by *S* at the time of input. Different retrieval models may be indicated if the information is not ordered, but unified, for example, in imaginal scenes.

REFERENCES

Atkinson, R.C. & Shiffrin, R.M. Human memory: A proposed system and its control processes. In K.W. Spence and J.T. Spence (Eds.) The psychology of learning and motivation: Advances in research and theory. Vol. II. New York: Academic Press, 1968.

Battig, W.F. & Montague, W.E. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. Journal of Experimental Psychology, 1969, 80, (3 Part 2).

Bower, G.H., Clark, M.C., Lesgold, A.M., & Winzenz, D. Hierarchical retrieval schemes in recall of categorized word lists. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 323-343.

Chandler, J.P. Subroutine STEPIT. Program QCPE 66, Quantum Chemistry Program exchange, Indiana University, 1965.

Clifton, C.E. & Gutschera, K.D. Evidence for hierarchical search processes in a recognition task. University of Massachusetts Tech Report 70-4, November, 1970.

Collins, A.M. & Quillian, M.R. Does category size affect categorization time? Journal of Verbal Learning and Verbal Behavior, 1970, 9, 432-438.

Feigenbaum, E.A. Information processing and memory. In D.A. Norman (Ed.) Models of human memory. New York: Academic Press, 1970.

Freedman J.L. & Loftus, E.F. Retrieval of words from long-term memory. Journal of Verbal Learning and Verbal Behavior, 1971, 10, 107-115.

Glanzer, M. Storage mechanisms in free recall. In G.H. Bower and J.T. Spence (Eds.) The psychology of learning and motivation: Advances in research and theory. Vol. V. New York: Academic Press, 1971.

Juola J.F., Fischler, I., Wood, C.I. & Atkinson, R.C. Recognition time for information stored in long-term memory. Perception and Psychophysics, 1971, 10, 8-14.

Landauer, T.K. Rate of implicit speech. Perceptual and Motor Skills, 1962, 15 646.

Landauer, T.K. & Freedman, J.L. Information retrieval from long-term memory. Journal of Verbal Learning and Verbal Behavior, 1968, 7, 291-295.

Meyer, D.E. On the representation and retrieval of stored semantic information. Cognitive Psychology, 1970, 1, 242-300.

Naus, M.J., Glucksberg, S., & Ornstein, P.A. Memory search strategy as a function of set-organization. Paper presented at annual meeting, Eastern Psychological Association, April, 1971.

Reitman, W. What does it take to remember? In D.A. Norman (Ed.) Models of human memory. New York: Academic Press. 1970.

Rundus D.L. An analysis of rehearsal processes in free recall. Technical Report No. 159, Institute for Mathematical Studies in the Social Sciences, Stanford University. August 21, 1970.

Seamon, J.G. Serial position effects in latency data: Search or strength? Submitted for publication, Journal of Experimental Psychology, 1972.

Segal, E.M. Hierarchical structure in free recall. Journal of Experimental Psychology, 1969, 80, 59-63.

Sternberg, S. High-speed scanning in human memory. Science, 1966, 153, 652-654.

Sternberg, S. Retrieval of contextual information from memory. Psychonomic Science. 1967, 8, 55-56.

Sternberg, S. Memory-scanning: Mental processes revealed by reaction time experiments. American Scientist, 1969, 57, 421-457.

Sternberg, S. Decomposing mental processes with reaction-time data. Invited Address. Presented at annual meeting, Midwestern Psychological Association, May, 1971.

APPENDIX A

Serial-Recall and Controlled Rehearsal Instructions

The present experiment will study information retrieval from long-term memory. In the first part of the experiment you will be presented with three groups of three words at a rate of 18 sec. per group - that is - you will be shown three words at a time with each three word group shown for 18 sec.

> S was shown an example of a three word group printed on an index card e.g., TREE - BUSH - FERN.

I want you to rehearse the group out-loud while it is being shown to you in time with the click that you hear. I want you to rehearse all three words between clicks for 18 sec. and to do this for each group. It is very important for you to rehearse consistently at all times and to think about the words that you are rehearsing.

> S was allowed to practice oral rehearsal with the sample word group.

What you did with the example, you will now do with three different groups of words. After the last group of three words has been shown and rehearsed, the word "READY" will appear on your screen. When the word "READY" appears you are to recall all nine list words by writing them on the sheet of paper before you. The first list word has to go in the first blank space, the second list word in the second blank space, etc., but the actual order in which you write the words on the paper is up to you.

> This point was clarified by example to insure that S understood the recall procedure.

You will have about 1 min. to recall the words from the time the word "READY" appears on your scrren. We will continue in this fashion until you have recalled the word list perfectly twice in succession. Any questions?

Remember, think about the words while you are rehearsing.

APPENDIX B

Memory Probe Instructions

In this part of the experiment I will test your memory of the word list you have just learned. You will now be presented with a fairly large number of test trials to which I would like you to respond as rapidly and as correctly as possible.

> S was shown an example of a five word category, a memory probe from the category, and the correct response, e.g.,
>
> Category DESK - TABLE - COUCH - CHAIR - LAMP
>
> Probe FURNITURE
>
> _ _ x _ _
>
> Correct Response "COUCH"

You will have test trials for all the words you have just learned and there will be more than one test trial for each word.

In order to time your answers I will use a timer which begins operating as soon as a test trial is flashed on the screen. By speaking loudly into the microphone you stop the timer and the computer will record your response and the time it took you to make the response. In using the microphone be careful not to make any noise before your actual response or you may trip the timer.

> S was shown how to use the microphone and cautioned about saying "Ahh..." before responding.

Lastly, do as well as you can on every trial. It is essential that we get accurate measurements on each word in the list. For your convenience there will be short rest pauses throughout the experiment, as well as several breaks during the session. Any questions?

Remember, respond rapidly and correctly.

APPENDIX C

Automatic - Recite Discrimination Instructions

I now want to change your task slightly by requiring a little more from you on each trial. In addition to giving me the correct response as quickly as you can each time, I want you to tell me how you recalled the response. If you are not aware of how you recalled it - you just knew it - that is, if the correct response was the only word that came to mind - then say the word "Automatic" after your response. If, however, the correct response was not the only word that came to mind - if you thought of one or more other words before making the correct response - then say the word "Recite" after your response.

> _S_ was encouraged to discuss this rule to insure that he understood how to make the Automatic-Recite discrimination.

It is very important that you understand and follow these instructions correctly. I want you to think about how you recalled a response only after you have made your response. Further, I want you to respond in whatever manner is natural for you. It is perfectly acceptable if all of your responses are Automatics or all Recites or some of each. What is important is that you tell the truth and do not guess. Remember, think about how you made your response only after you have made your response. If the correct response was the only word you thought of, say "Automatic". If you thought of one or more other words before making the correct response say "Recite". Any questions?

Respond quickly and accurately.

APPENDIX D

General Instructions

The first list was a practice list and is not an essential part of the experiment. Its function was to introduce you to the procedure. For the remainder of the experiment you will follow the same rules on each of seven new lists. The single exception is that a response of "Automatic" or "Recite" is required to the memory probes of all 20 trials, rather than only the last 10 trials as was done in the practice list.

During each daily session you will learn two lists, thus making a total of eight lists in four days. Each list you learn will contain new words, and you will never be reexamined on a list learned previously. Any questions?

APPENDIX E

Stimulus Materials

Category	Category Instances					
1 GEM	DIAMOND	RUBY	EMERALD	JADE	PEARL	OPAL
2 TIME	HOUR	MINUTE	WEEK	YEAR	DAY	MONTH
3 FAMILY	UNCLE	FATHER	SISTER	COUSIN	NIECE	SON
4 METAL	IRON	COPPER	STEEL	BRASS	TIN	ZINC
5 ANIMAL	DEER	WOLF	RABBIT	MOOSE	SKUNK	FOX
6 CLOTH	COTTON	WOOL	SILK	NYLON	VELVET	SATIN
7 COLOR	BLUE	RED	GREEN	YELLOW	ORANGE	BLACK
8 BODY	LEG	FOOT	NOSE	FINGER	NECK	HAND
9 FRUIT	APPLE	PEAR	BANANA	PEACH	GRAPE	CHERRY
10 HOME	TENT	COTTAGE	IGLOO	CABIN	MANSION	TRAILER
11 DRINK	BEER	GIN	WINE	VODKA	SCOTCH	RUM
12 JOB	LAWYER	TEACHER	NURSE	FARMER	BANKER	GROCER
13 WATER	RIVER	LAKE	OCEAN	STREAM	SEA	POND
14 WEATHER	SLEET	RAIN	WIND	HAIL	FOG	THUNDER
15 CLOTHES	SHIRT	SHOE	COAT	BELT	SCARF	VEST
16 MUSIC	PIANO	DRUM	VIOLIN	FLUTE	HARP	TUBA
17 BIRD	ROBIN	EAGLE	CROW	HAWK	PARROT	OWL
18 VEHICLE	CAR	BUS	TRAIN	TRUCK	WAGON	BIKE
19 INSECT	FLEA	ANT	SPIDER	BEETLE	WASP	MOTH
20 FLOWER	TULIP	DAISY	ORCHID	IRIS	LILAC	VIOLET
21 TREE	OAK	MAPLE	PINE	ELM	BIRCH	SPRUCE
22 FISH	TROUT	SHARK	PERCH	SALMON	MINNOW	TUNA
23 MEAT	BEEF	PORK	STEAK	VEAL	HAM	LIVER
24 EMOTION	LOVE	HATE	FEAR	ANGER	JOY	SORROW
25 DESSERT	CAKE	PIE	JELLO	PUDDING	COOKIE	SHERBET
26 READING	NOVEL	LETTER	ESSAY	POEM	JOURNAL	THESIS
27 TOOL	HAMMER	CHISEL	WRENCH	PLIERS	CROWBAR	RULER