72-14,840

DeCOSTER, Lee Frederick, 1945-
MEMORY CODES REVEALED BY RESPONSE
LATENCIES.

The University of Michigan, Ph.D., 1971
Psychology, experimental

MEMORY CODES REVEALED BY

RESPONSE LATENCIES

by

Lee Frederick DeCoster

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Psychology)
in The University of Michigan
1971

Doctoral Committee:

    Professor Irwin Pollack, Chairman
    Professor Wilfred M. Kincaid
    Professor Arthur W. Melton
    Professor Daniel J. Weintraub
    Dr. Sylvan Kornblum, Mental Health Research Institute

DEDICATION

This work, as the culmination of my years in college and graduate school, is dedicated to my parents, Francis and Hilda DeCoster, in appreciation of the steadfast love and nurturance which they never failed to provide in times of celebration and times of despair; and to my sons, Barton Trent and Robb Frederick, as a source of encouragement that they also pursue those goals toward which they are inclined.

In keeping with this dedication, my writing has sought to minimize the use of technical jargon and the discussion of cryptic theoretical issues. A dissertation is a scholarly work. I believe the highest form of scholarship accepts the responsibility of providing succinct communication for the interested layman. To the best of my ability, I pursue that goal.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

(

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER I

THE PROBLEM

Scientific endeavors can be partitioned into two types of activities: generating questions and developing methodologies. The inductive process of constructing theories and the deductive process of formulating predictions are inherent aspects of the former activity. When incompatible predictions are derived from separate theories the scientist has identified an issue. The resolution of issues requires sound methodology for the proper comparison of theories.

Psychology, as a science, inherits the task of questioning the nature of mental operations. That cognitive events are not observable, necessarily being inferred from measurable events, places a heavy burden on the inductive process. It is this extreme complexity of scientific inference involved in unveiling mental processes which prompts psychologists to explore and perfect new methods and experimental paradigms.

Recently, Saul Sternberg has developed both an experimental paradigm (Sternberg, 1966) and a rational-logical theory for operationally differentiating various cognitive processes (Sternberg, 1969; 1971). The present study attempts to apply these procedures and tools of inference, which were developed in another context, to a currently debated theoretical issue. The method examines response latencies in a memory-dependent, two-choice classification task. These time measures should be helpful in resolving the question: "What are the functional properties of those memory traces used in short-term recognition memory?" While the attributes of various types of memory have been investigated through a wide range of procedures, the form of the memory trace used in this type of task has not been determined. A clarification of the theoretical issue concerning memory precedes an elaboration of the methodology.

THE ISSUE

What Form Memory?

Introspection leads us to believe that words are stored in memory with large amounts of retrievable information. When asked, we can usually provide correct answers to questions about the spelling, pronunciation, overall shape, syntactical classification, and various aspects of the meaning of thousands of words. However, further reflection suggests that many things we could report accurately are not immediately available, but come after some special mental activities. For example, if asked how many letters are in the word "little," we usually need to first spell out the word and then count each letter. Moreover, we often have gaps in the usual array of memorized attributes of a word. Most of us need more than a second's time to recognize that "table" can be a verb, yet few adults require half that time to pronounce the word. Often a common word will not be immediately recognized when it appears in all capital letters, such as STIR, or POESY; conversely, what is the meaning of "i"? As a youngster I understood the meaning of "epitome" for years before I connected it with the spelled form, although I had seen it written many times. Many of us have a vocabulary of ethnic words which we understand and use, although we have never knowingly seen them in print. The word is in memory, but not in all of its forms.

While we may not wish to put much faith in such anecdotal evidence, we should not ignore its value in calling to our attention the fact that some properties of memory are immediately available while others require more time. Since we wish to know the "energy form" (Posner, 1967) or the functional code by which words are retained in memory, we must take a position regarding (1) the interpretation of variations in the time required to answer questions about the properties of words we have memorized, and (2) the type of memory we are prepared to investigate.

First let us consider that we observe a reliable difference in the time needed to correctly report one property of a word (e.g., its pronunciation) relative to some other property of the same word (e.g., its meaning).

We can interpret this finding in at least three ways: (1) There is a single memory code for the word, from which we can derive other features by performing additional cognitive operations. For example, if memory were auditory, we might pronounce the word, and then determine which letters were needed to produce such a sound by applying linguistic rules. We would then report the visual form of the word as calculated from its pronunciation. (2) There are several separate memory traces, all specifying different properties of the same word. However, it merely takes longer to either retrieve or to "decode" some types of these traces than other types. Thus we might have a visual memory trace and a set of semantic memories (synonyms), but we may be faster at locating the visual traces. (3) There are several memory traces for each word, all of them being chained together in such a way that we must search the chain in a fixed sequential order. For example, the chain might always first supply the spelling or visual form, then the pronunciation or auditory form, and finally the meanings of the word. While these three possibilities do not provide an exhaustive account of models of memory, they serve to give us a feel for the complexity of interpreting the hypothesized differences in response latencies. Throughout this thesis we shall employ the first model, which assumes (in an admittedly naive manner) that additional time to respond implies additional processing. The "more time, more processing" approach will be explained in greater detail in the later section concerning methodology. Although we might generate elaborate theories to deal with our topic, it will be to our advantage to construct the simplest possible model, modifying it only when the evidence is largely incompatible with the model's predictions.

The specification of the type of memory to be investigated is a more formidable task. In 1958, Donald Broadbent proposed that there were at least two structurally distinct types of memory: a short-term store which held information for time intervals "of the order of seconds," and a more permanent or long-term store. Since that time, there has been a proliferation of titles for different forms or different stages of memory operations. Extreme positions have been recently stated by Murdock (1970), who claims

there is no need to distinguish types of memory, and by Wickelgren (1970), who argues that there is empirical justification for at least four stages of memory. These stages have been differentiated with respect to such factors as their means of acquisition, the amount of information held in store, processes of consolidation and decay, and the operations required for retrieval.

Donald Norman (1968) argues that there is one stage, primary or short-term memory, which involves an auditory or acoustical representation, whereas a second stage, secondary or long-term memory, is thought to involve some type of semantic representation. The basis for this inference relied upon (1) an analysis of errors in recall or recognition using confusion matrices (e.g., Conrad, 1964, 1967), and (2) upon the amount of interference produced by stimulus materials varying in similarity with the to-be-memorized item (e.g., Dale & Gregory, 1966; Posner & Konick, 1966; Wickelgren, 1965). The similarity may be judged according to various dimensions, such as visual, articulatory, phonemic, or semantic similarity. Recently these two procedures have come into question due to their imprecision in discriminating among different forms of representations (Wickelgren, 1969), and due to their lack of agreement with results obtained using other methodologies (Shulman, 1970; Wickens, 1970).

We shall reverse Norman's approach. Rather than use evidence regarding memory codes to identify different types of memory, we shall leave open the theoretical specification of the type of memory under consideration in order to concentrate on how we might obtain evidence to clarify the nature of memory codes. Norman's "facts" become our issue. We shall take the objective position that we are studying whatever type of memory is required by our task. When we describe the task as a short-term recognition task, we mean by "short-term" only that the test comes less than a minute after presentation of the to-be-memorized words. By "recognition" we mean the subject is shown a set of items, some being the same as the memorized item, others being different. The subject must state (recognize) upon presentation of each item whether it was the same or different from the memorized item. It is in this sense that we ask, "What are the functional codes used in short-term recognition memory for verbal items?"

Three Hypotheses

The results of various lines of research enable us to formulate
three hypothesized forms of the memory trace. In keeping with our emphasis
on simplicity, we shall consider them as mutually exclusive possibilities
within a single trace, although it should be clear that combinations of
functional properties are possible (e.g., multi-dimensional trace theories
have been proposed by Bower, 1967, and by Norman & Rumelhart, 1970).
The Auditory Hypothesis proposes that a memory trace preserves only the
sound features of a word, being isomorphic with the perception of spoken
words. The Semantic Hypothesis claims that a memory trace retains the
meaning of a word. Finally, the Visual Hypothesis states that a memory trace
is isomorphic with the physical-spatial form of written words. Each hypo-
thesis will consider that a word represents an auditory, semantic, or
visual unit, although each unit may have subunits or features. Sample
evidence related to each hypothesis will be considered in order.

The principal evidence for the Auditory Hypothesis has come from
error analyses. Confusion matrices can be understood through comparisons
between different letter pairs. Consider that we briefly flash the
letter "B" on a screen and a subject erroneously reports the letter "P."
We are not at liberty to determine whether this error resulted from
the similarity in the visual form of the two letters, the similarity
in their sound, or their common means of articulation. However,
confusing an "E" for an "F" might incline us to attribute the error
to visual similarity, since the letters are not articulated alike,
nor do they have the same physical sound pattern. By the same reason-
ing, confusing a "B" with any of the following letters -- C, D, G, T, V,
or Z -- would be attributed to the acoustic similarity (same trailing
phoneme) rather than visual similarity. This implies that the letters
had an auditory code and not a visual code.

Wayne Wickelgren (1969) pointed to the difficulty in differentiating
among acoustic hypotheses by asking, does the subject encode the way an
item sounds (auditory) or does he store that set of muscle commands re-
quired to pronounce the item (articulatory form)? Unfortunately, these

two possibilities, while conceptually distinct, are naturally confounded for English letter names. Although unmentioned by Wickelgren, we should further note that the physical wave form of sound is correlated with auditory perception. In general, letters which are perceived as distinct have dissimilar patterns of speech production and have greater physical differences in spectral frequencies. As an example of similarity in perception and articulation, we can consider the pairs B-D, M-N, and D-T. These pairs sound similar (perception) and have similar articulatory features. The articulatory difference between B and D, and between M and N is simply a difference in the "place" of articulation, while the difference between D and T involves a difference in "voicing." In general, the more dimensions of articulation which differ, the greater the difference in both perception and wave form. To avoid theoretical issues of speech perception, we will avoid differentiating these hypotheses, using the terms auditory code and acoustic form of memory to refer to the same general hypothesis, their interpretation being that the memory trace is isomorphic with the perception of a spoken word.

Interference studies typically give subjects a list of words and provide a test of recognition, or recall, immediately thereafter. By constructing lists containing homonyms, synonyms, or unrelated words, comparisons of accuracy in recall are possible. For example, when this procedure was used by Kintsch and Buschke (1969), using 16-word lists followed by single-word cued recall, the authors found less accurate recall of early items in the synonym list, relative to a control. Accuracy was less for later items with the list containing homonyms. These performance decrements relative to a control were attributed to interference, presumably produced by similarity between memory traces. These data suggest that the memory of words which have been available for longer time intervals contain attributes of meaning, while memory traces available for shorter time periods are primarily auditory.

Another procedure involving interference provides further evidence of semantic properties in short-term memory. Delos Wickens (1970) reports that recall of word triads was near perfect for the very first trial, but that accuracy dropped considerably on the second and

third trials. By the fourth trial performance was very near an asymptote
(plateau) of less than 60% correct recall. Although additional trials
with the same types of words continued to show this low level of accuracy,
it was possible to have the subject's performance return to its initial,
near-perfect level. This was done by changing the stimulus materials.
Wickens argues that the demonstration of a return to a high level of
performance implies that any drop in accuracy must have been due to
interference. Once again, it is assumed that this proactive inhibition
was produced by similarity between memory traces. Changing the critical
stimulus dimension means the new materials are functionally dissimilar,
hence the source of interference has been removed. By this reasoning,
those dimensions which, when altered, fail to eliminate interference must
not be represented in the memory trace. They are irrelevant to the
memory code. Conversely, release from proactive inhibition signifies
that the altered dimension is a critical component of the trace.
Wickens found that changing physical features, such as the color or
size of the words had little effect. Altering the taxonomic category
of the words, as in going from names of flowers to names of trees,
substantially improved performance. Therefore these semantic categories
must be represented in the memory code. Consequently, Wickens claims
evidence for the Semantic Hypothesis.

Additional evidence for semantic encoding has been supplied by
Harvey Shulman (1970) based on an analysis of reaction times. However,
since this procedure is an outgrowth of previous reaction time research
conducted by Posner and by Cohen, it will be instructive to digress from
the chain of evidence supporting the Semantic Hypothesis in order to
consider the earlier research.

Concern with the temporal properties of word and letter recogni-
tion springs largely from experiments using Sperling's (1960) procedure
wherein large arrays of letters are presented visually for a brief dura-
tion followed by a cue to recall a subset of the array. Such studies have
provided evidence of confusions related to the visual form of the stimuli
(Chase & Posner, 1965; Keele & Chase, 1967). Sperling (1963, 1967)
interprets the difference in the types of errors resulting from this

task, relative to others using longer exposures and delayed recall, as
evidence that there is a visual form of memory which holds information
until it can be transformed into an auditory representation.

Support for this interpretation has come from reaction time
experiments by Michael Posner and his colleagues (Posner & Keele, 1967;
Posner & Mitchel, 1967; Posner, Boies, Eichelman, & Taylor, 1969).
The procedure in these experiments involved the sequential presentation
of a pair of letters, only one letter being visible at a time. A decision
was required after each pair. Typically, the first letter was presented
to the subject in either upper case (e.g., "A") or lower case print
("a"). The instructions were to respond "Same" if the second test
letter had the same name as the first letter, otherwise respond
"Different." The term "name" referred to the way a letter was pro-
nounced. The reaction time was quicker when the pair of items were physi-
cally identical, relative to being similar only in name. The advantage
of the identity match was greatest when the interval between the removal
of the stimulus standard and the onset of the test letter was short.
The non-identical name match was at least as rapid for inter-stimulus
intervals greater than 1.5-2.0 seconds. Thus the Visual Hypothesis
finds support in studies where the memory trace must be used just briefly
after it has been established.

Gillian Cohen (1968) made use of another of Posner's procedures --
the simultaneous presentation of letter pairs. Cohen used word pairs,
rather than sets of letters. The pairs were either identical (sew-sew),
homonyms (sew-so), or synonyms (sew-stitch). Each subject made a Same
vs. Different judgment to each pair by depressing response keys using
separate hands. The basis for the response classification was one
of three instructional rules, all subjects working with each instruction.
These rules required a Same judgment if the words were (1) visually
similar, (2) acoustically similar, or (3) semantically similar, as
in the above examples.

Perhaps the most interesting result involves the reaction time
to identity pairs when the various instructions were in effect. The
Same response required 678 msec. under visual match instructions, 716

msec. for the acoustic match, and 839 msec. under the semantic rule.
While the author failed to interpret the results in this manner, it
is not unreasonable to assume that the difference between the visual and
acoustic reaction times represents the time required to translate a
visual image into an auditory form (i.e., 38 msec.) while the semantic
minus visual time represents the time for translation into a semantic
unit (161 msec.). When two distinct visual patterns are presented and
the words are either auditory or semantic equivalents, then the response
time should reflect the time for two translations. This model predicts
a total time for acoustic equivalents of 678 + 2(38) = 754 msec. Un-
fortunately this predicted value is not even near the obtained time of
839 msec. For semantic equivalents the predicted time would be
678 + 2(161) = 1000 msec. This time estimate is right on the obtained
value of 1001 msec.

The Different judgments produced reaction times which were
highly similar in magnitude and variations to the Same responses. The
only statistically significant effect was that subjects using the visual
match rule were slower to make their Different judgments to homonym pairs
than to either synonyms or unrelated words. Both unrelated words and
synonyms averaged 612 msec. for a Different judgment, compared to 725 msec.
for homonyms. This finding suggests auditory properties were present in
the memory code, interfering with the classification decision.

Further evidence for auditory encoding of the items came from Cohen's
analysis of errors. Subjects made about half as many errors when working
with the acoustic rule. Moreover, the greatest number of errors by
word-type involved false positive judgments to homonyms. The acoustic
properties of the items were apparently functioning while the subject was
attempting to make both visual matches and semantic judgments.

Harvey Shulman (1970) combined the measurement of reaction times
under rules for semantic and acoustic similarity with the list-plus-probe
technique, discussed above. Shulman's procedure called for the sequen-
tial presentation of 10 words, followed by a matching rule (acoustic vs.
semantic similarity), and finally by a recognition probe which could be
(1) identical to a list item, (2) a homonym, (3) a synonym, or (4) un-

related to list items. The subject was to respond Yes (i.e. the item is equivalent to one in the list) or No to the probe.

Shulman had hoped that the rule uncertainty, produced by delaying the instruction until after the list had been presented, would force semantic encoding to the degree that such encoding was possible. He further reasoned that semantic encoding (setting up a memory trace with semantic properties) may require more time to establish relative to an auditory trace. Consequently, the rate of word presentation was varied, having single-item exposure durations of 250, 600, or 1300 msec. Shulman predicted that the response time to synonyms would be shorter with slower rates of presentation.

The results showed equal proportions of correct Yes responses for identity items and homonyms, but fewer correct synonym judgments. Reaction times were quickest for identity items, next for homonyms, and slowest for synonyms. All three functions produced similar patterns of response times as a function of the position in the list of the matched item. The average difference in Yes responses was about 165 msec. between identity items and homonyms, and approximately 255 msec. between homonyms and synonyms. While the reaction times increased for all types of items with slower rates of presentation (contrary to the prediction for synonyms) the proportion of false positives decreased for synonyms with slower presentations while increasing for homonyms. Although Shulman interpreted both reaction time and error data as evidence for semantic encoding, it is only this last finding which can not be explained solely from the Acoustic Hypothesis (assuming the sound is retrieved and the meaning then derived from some long-term store). However, since the error rate for synonyms was always greater than for homonyms, even this interaction of errors by rate of presentation is unconvincing. Additionally, we should note that greater speed and fewer errors for probes identical to test items supports the Visual Hypothesis.

The issue remains: Is short-term memory essentially visual, auditory, or semantic? The bulk of evidence supports the Auditory Hypothesis, but this model is unable to account for faster responding to visually-similar materials. Those supporting the Visual Hypothesis

can not explain the acoustic similarity among errors, nor the semantic interference. The Semantic Hypothesis is damaged by those studies using response times and error confusions, finding its sole support in those studies which rely on the concept of interference.

THE METHODOLOGY

A Conceptual Framework

Saul Sternberg has developed a methodology which provides considerable leverage in resolving theoretical issues involving memory. Since this method requires a theory for interpreting reaction time data, it will be useful to briefly review the principles of Donder's Subtraction Method, which has already been applied to Cohen's data. Consider that a subject is seated before a display screen with instructions to press a telegraph key when a light appears. At the moment we initiate the light signal, we start a clock, stopping it at the time of the key press. Let us interpret this time period as representing the sum of several shorter time intervals, each period corresponding to the time needed for some type of processing of the stimulus information. For example, we might imagine it takes time for the light energy to get translated to a neural message, more time for a decision that this message is sufficient to warrant a response, and additional time to signal the appropriate muscles to act, resulting in the measured key press. We are assuming that the decision process can not start until it receives some message and further that the muscle command is contingent upon the decision process. That is, we conceive of separate functionally-defined processing centers which operate one-at-a-time, or sequentially.

If we change the experiment such that the subject is correct in responding only when one type of stimulus appears, as only a red light, other colored lights being ignored, we would discover that the subject is slower in correctly pressing the key. Our method implies that some additional processing was necessitated by the new task requirements.

Donders proposed that the magnitude of the difference in response times
(comparing the two situations) represents the time needed for additional
stages of processing. This Subtraction Method assumes "pure insertion"
of a new stage of processing. That is, the slower reaction times which
result from increased task requirements are produced by the addition of
functioning centers, rather than increments in the time needed at each
existing stage of processing. Technically, these stages must be stochas-
tically independent.

Sternberg (1971) argues that the strong theoretical assumptions
of the Subtraction Method are not necessary. He proposes an Additive-
Factors Method that also requires the assumption of sequential processing,
but provides that an experimental variable can affect more than one
stage. It is possible to determine which experimental manipulations
affect which inferred stages of mental processing by examining the data.
Data need to be statistically evaluated for main effects and interactions
in order to make the necessary inferences. This model does not suggest
that processing centers operate in an all-or-none fashion. Even with
these less restricting assumptions, the method is capable of establishing
both the number of separate functional (inferred) stages which are operat-
ing, and the time parameters for any given stage. A more complete account
of this theoretical framework for interpreting reaction time data can be
found in Sternberg's works (1969, 1971).

The Experimental Paradigm

Sternberg (1966) also developed a particular experimental paradigm
which helps specify a crucial feature of memory (viz., a search to deter-
mine the presence or absence of a code). The procedure calls for presenting
a list of from one to six items to a subject for memorization. The subject
can be given a few seconds or several minutes to memorize the items which
can be presented either sequentially or simultaneously. Then a single-item
test is displayed visually. The subject's task is to either press
"Yes," signifying that the item is the same as one of the memorized
items, or "No," that it is different. Next the procedure may call for

either a new list of to-be-memorized items (thereby having only one test per trial), or several more tests can be given for the same memorized items. The to-be-memorized items are said to form the Positive Set since they require a Yes response. The dependent variable of greatest interest has been the response latency, the time measured from the onset of the test item to the pressing of a response key. Separate hands are used to indicate a Yes vs. No response.

Typical results show a mean response time of just over 400 msec. for both Yes and No responses when a single digit or letter has been memorized. The Yes time is typically faster when Yes and No responses are equally likely. Increasing the probability of either response will lower its average latency relative to the alternative response. The average reaction time increases in equal increments of 35-40 msec. as the number of items in the Positive Set increases. Consequently, the easiest way to describe the data is to plot reaction time as a function of the size of the Positive Set, in which case a straight line having a slope of 35-40 msec. per item connects all observations. If the lines are fitted separately to all Yes and all No responses, the lines will be nearly parallel, the No line intercepting the ordinate slightly above the Yes line. The interpretation of such data is quite straightforward, although the theoretical argument is detailed. It appears that a subject recognizes the test item and then compares this item with each-and-every item in memory. A Yes response is initiated if the test item matched one of the memorized items; a No response being made in the absence of a match. Thus the subject makes a serial, exhaustive search of the Positive Set.

The theoretical argument considers three questions, each issue having two possibilities. They are: (1) Are the comparisons between the test item and the items in memory made one-at-a-time (serially) or simultaneously (in parallel)? (2) On correct Yes tests, is the search ended as soon as a match is made (self-terminating) or are all possible comparisons made before the response is initiated (exhaustive)? (3) Is the length of time to make a single comparison a constant, or is it variable (zero vs. non-zero variance)? A fourth, contingent issue is:

When there is more than one test after a list has been memorized, there are possibilities of both response and stimulus repetitions. Consider that there are two items in the Positive Set, the digits "1" and "2." Presenting a "3" as the first test item results in a No response. A "2" as the second test item requires a Yes response. Since this second response is a different one, it will be termed a Non-Repetition. The change could have gone from a Yes first response to a No, in which case we would classify the second response as a No Non-Repetition. If on the third test a "1" is presented and the subject responds Yes again, we have a response repetition. However, the stimulus has changed, constituting a stimulus non-repetition. Following Bertelson (1965), this form of response repetition will be termed an Equivalence response. Presenting another "1" as the fourth test would involve both a response repetition and a stimulus repetition, which we will call an Identity response. It is possible to have Identity or Equivalence responses for both Yes and No classifications provided the Positive and Negative Sets contain more than one item (with one item a Yes response repetition must be an Identity).

If a subject always searches through the entire Positive Set before initiating a response, there is no reason to expect a difference in the reaction time among the repetition categories. However, Sylvan Kornblum (1969) has shown that sequential effects can be an important determinant of response latencies, affecting a variety of tasks. Repetition responses require less time than non-repetitions, and consecutive repetitions show a continual reduction in latencies. What would be the interpretation of a main effect of repetitions in the Sternberg paradigm? This is where the Additive-Factors Method is needed. The method states that an experimental variable producing a main effect, but no interaction with other variables, influences a separate stage of processing relative to the other variables. Thus if repetitions have a lower intercept than non-repetitions, but have the same slope as a function of the size of the Positive Set, it is safe to assume that this variable affects some stage of processing other than the memory-comparison stage. Presumably the effects would be due to a facilitation in some process which controls the muscle action -- a response process.

If response repetitions, either Equivalence or Identity responses, have a different slope than the Non-Repetition responses, we can not conclude that the memory-comparison stage is unaffected. Such evidence would be damaging to the serial exhaustive search model. The present study will examine latencies for each of the three categories of sequential responding.

APPLYING THE PARADIGM

As stated previously, the present study attempts to apply the Sternberg memory search paradigm to the question: "What is the functional code used in short-term recognition memory?" The present study employs three instructions and three classes of stimuli. The instructions require that the subject compare the test item with the items in memory where a match is defined as locating a memory trace which is (1) visually identical, (2) acoustically identical, or (3) semantically the same as the test item. Thus the instructions state that the test item either (1) "looks the same," (2) "sounds the same," or (3) "means the same" as an item in memory in order to receive a Yes response. The stimulus materials are either homonyms, synonyms, or unrelated items. By using the same subjects and the same stimulus materials with each of the instructions, the relative efficiency (i.e., speed) of the memory searches under the different rules can be assessed.

The rationale is quite simple, provided we adhere to our three original hypotheses. Our basic assumption is that the Sternberg paradigm does reflect a process which functions similar to the serial exhaustive search model. If this is true, then the slope corresponds to the time required to make a single memory comparison. The basic prediction is that the data will show maximum predictability and the shallowest slopes when the units representing the size of the Positive Set are the same as the form of the memory trace. That is, if the Auditory Hypothesis is correct, the slope will be flattest when the size of the Positive Set is defined by the number of separate sounds, rather than the number of distinct meanings of physically different visual patterns.

At this point we need to clarify the concept of the "Positive Set." This term can be used in two ways. (1) The Empirical Positive Set refers to those visually distinct items which would require a Yes response if displayed as a test item on a given trial; this set changes with each trial as the to-be-memorized items change. (2) The Theoretical Positive Set refers to the inferred memory traces. For the Visual Hypothesis these sets are the same since each distinct item will have a separate memory trace. For the Auditory and Semantic Hypotheses these sets may differ. For example, if the to-be-memorized set contains the items "1, ONE," and the instruction is "looks the same," the Empirical Positive Set involves two items, since two separate displays require a Yes response. The Theoretical Positive Set contains only one unit under both hypotheses, since only one sound or one semantic unit is included. Unless noted otherwise, the term Positive Set can be interpreted as the Empirical Positive Set.

A description of Blake Wattenbarger's research will illustrate the type of derivations permitted by our basic model. The elaborated rationale is developed in Chapter III.

Wattenbarger (1970) used the Sternberg paradigm in order to extend Posner's comparisons of physical identity vs. name matching. These two types of matching are similar to the Looks and Sounds Conditions, respectively, of the present study. We recall that Posner found faster response times to physically identical pairs of letters provided that the inter-stimulus interval was less than about two seconds. Wattenbarger's study involved three conditions: Name Identity, Physical Identity, and a Control. For Name Identity, letters had to have only the same name to receive a Yes response, although they could be in either upper or lower case regardless of the case of the memorized letter. For Physical Identity, letters had to have both the same name and the same case as the memorized letters (the to-be-memorized set would contain a mixture of cases wherever possible). The Control presented all letters in the same case, such that test letters with the same name automatically were in the same case. As in the present study, Wattenbarger sought to compare slopes between conditions in order to infer the form of the memory trace.

A simplified statement of Wattenbarger's rationale attends to the difference between the slopes for the Physical Identity vs. Name Identity Conditions. If memory involves visual images, twice as many comparisons would be necessary for the Name Identity Condition relative to the Physical Identity Condition since, in the former, subjects would need to store both the upper and lower case version of each letter. Consequently, the slope for the Name Identity Condition, when plotted as the number of distinct letter names, would be twice as steep as that of the Physical Identity Condition, where the slope is plotted as the number of physically distinct visual patterns. On the other hand, if memory codes are auditory, the shallowest slope will be in the Name Identity Condition. In the Physical Identity Condition the subject will presumably store both the letter name and some additional code denoting which case (upper or lower) is correct. If this additional code requires a comparison time equal to that for the letter name, the slope of the Physical Identity Condition will simply be twice as steep as that of the Name Identity Condition.

The obtained slopes, combining Yes and No responses, were 97 msec. per character for Physical Identity vs. 57 msec. per letter name for Name Identity. Based on the above reasoning, these data support the position that memory codes are names, or in the terminology of the present study, auditory codes. The Control Condition slopes were 54 and 44 msec. per letter for upper and lower case characters, respectively. The similarity of these slopes to that of the Name Identity Condition is consistent with the conclusion that name or auditory codes were used throughout. However, contrary to our Auditory Hypothesis, a breakdown of the No responses in the Physical Identity Condition revealed a steeper slope (102 vs. 89 msec. per character) and higher intercept (816 vs. 555 msec.) for those items having the same name but a different case, relative to items differing in both name and case. Finally, the error rate was lower for the Name Identity Condition relative to the Physical Identity Condition, 3.1% vs. 6.0%, again supporting the Auditory Hypothesis.

The data from Wattenbarger's study are less than satisfying for two reasons: (1) the range of response times, for both slopes and intercepts, departs from other studies, and (2) the data provide no consistent

clue as to what the subject was doing in the Physical Identity Condition
The 57-msec. slope of the Name Identity Condition is considerably greater
than the 35- to 40-msec. slopes obtained by Sternberg (1969) using digits.
Also, Sternberg's intercepts were typically below 400 msec., compared
to the 528- and 630-msec. intercepts of Wattenbarger's two principal
conditions. The Auditory Hypothesis, as applied to Wattenbarger's design,
predicts (1) the intercepts for the Physical Identity and Name Identity
Conditions should be identical, and (2) there should be no difference
between slopes or intercepts for the two types of No response in the
Physical Identity Condition. Neither prediction was confirmed.

An alternative explanation of Wattenbarger's findings assumes the
Physical Identity Condition elicited a different form of memory trace
(viz., visual) which required more time for single comparisons. But this
model also makes the unverified prediction of no difference between the
two types of items requiring a No response (differ by case alone vs.
differ by both case and name). At best, Wattenbarger's results suggest
that a memory search process can be conducted most efficiently through
auditory memory traces. Whether or not other forms of memory trace were
used can not be determined.

ORGANIZATION OF UPCOMING CHAPTERS

Now that the issue and methodology which prompt this thesis have
been outlined, we are prepared to consider the procedures and results of
the present research. Six chapters describe the procedures, report the
results, and discuss the implications of three experiments. Chapter II
provides (1) an overview of the experimental design, (2) an account of the
assignment of subjects to conditions, and (3) a description of the equip-
ment and procedures of Experiment I. Since the procedures of the other
two experiments were highly similar to those of the first experiment,
any modifications will be explained as each additional experiment is in-
troduced. Chapter III reports the findings of Experiment I for both
response latencies and errors. The fourth chapter discusses puzzling
results of Experiment I, describes the rationale and results from Experi-

ments II and III, and outlines those aspects of the data requiring clari-
fication. Chapter V resolves most of the remaining complications by
partitioning the data according to sequential patterns of responding.
A final analysis of the results from Experiment I takes place in Chapter
VI, preceding the conclusions of the final chapter. In general, the
presentation of results and the accompanying discussions are organized
about a single theme -- what is the functional code used in short-term
recognition memory?

CHAPTER II

OVERVIEW OF EXPERIMENTATION

Design

Three experiments were needed to satisfactorily resolve the
basic issue. Experiments I and II had a common design. Both involved
the use of three experimental conditions, all subjects serving in every
condition. All possible orders of serving in the three conditions could
be incorporated by using six subjects, each receiving a different se-
quence of instructions (i.e., 1-2-3, 1-3-2, 2-1-3, etc.). Experiment I
used one six-subject block whereas Experiment II used two such blocks
(12 subjects).

Each of the three conditions required a one-hour testing session.
The term _series_ refers to a set of three sessions wherein each of the
three conditions was represented once. Experiment II used a single
series, whereas Experiment I involved three series. That is, each of
the six subjects in Experiment I participated in each of the three con-
ditions three times, amounting to a total of nine one-hour testing
sessions (these subjects also had a one-hour practice session, bringing
the total to 10 hours). No subject worked more than one session on a
given day.

Different subjects were used for each experiment, except that
three subjects from Experiment I served in Experiment II for four sessions
at a single condition. Their participation was an addition to the above-
described design, incorporated to obtain a more stable estimate of
performance in this condition. Also, an additional subject for each
condition worked for a single session in Experiment II. This was done
in order to provide a more reliable estimate of first session performance
since these responses would be uncontaminated by any possible influence of

21

serving in other conditions. Thus a total of five subjects worked in each condition as their first session in Experiment II.

Experiment III involved six subjects serving one session in the only condition in that study. One subject was subsequently eliminated, reducing the total to five subjects. Since this condition called for subjects to attempt to eliminate the naming of words, it seemed unwise to have these subjects participate in these conditions which required that the to-be-memorized list be read aloud and actively rehearsed.

Subjects

A total of 18 females and 11 males, all enrolled at the University of Michigan, began as subjects, with the data from 17 females and 9 males being reported. The female:male breakdown by experiment was 5:1, 10:5, and 2:3 for Experiments I, II, and III, respectively. All but two subjects were right-handed, with one additional subject reporting he was ambidextrous, although he wrote consistently with his right hand.

The subjects in Experiment I were all experienced, working more than four hours per week in a computerized psychophysics laboratory. Each was paid $2 per hour, plus a bonus of $2 for completing the 10 sessions. The first session for each consisted of instruction and . practice at all three conditions of the experiment. Thereafter these subjects set up the equipment and worked alone.

Naive subjects served their first hour in Experiment II to fulfill a psychology course requirement. After completing that hour, each was asked to return for two additional sessions at $2 per session. All but one of the subjects volunteered to return for pay. The data from the non-volunteer (who expressed the wish to return, but an inability to spare the time with several exams scheduled at the time for experimentation), and that from two other participants filled the need for three one-session-only subjects. The experimenter was present only during the seven practice trials in each session during Experiment II. The subjects worked alone for the test trials.

Subjects in Experiment III were obtained from the Psychology Department Subject Pool in the same manner and at the same time as those in Experiment II, but these subjects were not asked to return for pay. Those selected for this experiment were in no apparent way different from those serving in Experiment II. The order of assignment to each experiment had been determined in advance. However, one subject produced such long latencies (the median being over one second), and expressed such a negative attitude toward the research, that her data was discarded prior to analysis.

In addition to the above subject, two subjects in Experiment I were dismissed -- the one being unable to serve the total of 10 sessions (one practice, nine testing sessions), the other being asked to stop when he was unable to reduce his error race below 5% after three sessions. Partial analysis of the data from both subjects revealed no major discrepancies from the results of the other subjects. Nevertheless, the a priori criterion of dealing with low error rate subjects was administered.

Equipment and Parameters

The stimulus arrays, stimulus parameters, and response measurements were constant throughout all experimentation. The visual display was presented on a modified Tektronix Model 611 storage display oscilloscope, controlled by a Digital Equipment Corporation Model PDP-8I computer. The letters were all capitals plotted upon a potential 9 x 13 grid with letter size being approximately 3.7 x 5.4 mm. With the subject seated approximately one meter from the scope, a three-letter word would subtend 0.8° of arc.

The subject indicated his decision by pressing one of two microswitch response keys, mounted on a table in front of the scope. A 5- to 6-ounce force was sufficient to trigger the switch, depressing the key 1/8 inch. The subject was free to use either his index or second ("middle") finger, one finger for each hand. The hand used in writing was designated for the "Yes" response. All responses were timed on the computer's counter, accurate to ±1 millisecond.

A _trial_ was defined as the sequence of (1) an _instruction_,
(2) a _to-be-memorized_ _list_, and (3) a series of _test_ items, presented
one at a time. The instruction was always a two-line display, reading
"INSTRX:" on the first line, and "SOUNDS THE SAME," or whatever instruc-
tion defined the condition, for the second line. The instruction remained
for two seconds, followed by a one-second blank period, and then the stimu-
lus array. The to-be-memorized list of stimuli was arranged vertically,
the top-most item centered on the screen. The remaining items were listed
below, centered according to the rule of adjusting one place to the left
for every two letters present (thus one- and two-letter items started
at location zero, three- and four-letter words starting at position -1,
five- and six-letter words at position -2). The duration of this display
was 1667 msec. + (d)333 msec., where d = the number of displayed stimuli.
A three-second blank period followed, before a test item appeared, always
centered. The test item remained on the screen until the subject responded.
However, the time was not recorded if it extended beyond two seconds
(this was extremely rare, even during the practice sessions).

Upon a response, a two-second blank period elapsed before the
next test item appeared. Trials for which data were evaluated always
involved 10 tests per trial, but practice trials had as few as four
tests per trial, varying between experiments. Two seconds was the minimum
time following the end of one trial before the instructions appeared for
the next trial. This time interval varied, since a slow tape reader had
to supply a variable length list of parameters and stimuli for the upcoming
trial following the last response to the previous trial.

While the stimuli for a given trial were determined by a parameter
tape, the sequence and content of test items was controlled by the computer
via the following rules: There were always two permanent word pools:
(1) the 12 potential stimuli, any of which might form the _Positive_ _Set_
of memorized stimuli on a given trial, and (2) a _Null_ _Set_ of 12 items
which could never appear as stimuli, and hence never could be correctly
classified with a Yes response. The latter items did appear as tests,
always requiring a No response.

Figure 1 illustrates both the oscilloscope displays and the
computer's organization of word pools. The probability of displaying
a test item which belonged to the Positive Set was always 0.50. These
items required a Yes response. A random selection determined (1) which
word pool was consulted and (2) given a pool, which item was to be
displayed as a test. All selections were made with replacement and
the probabilities had no string-length constraints. A new Positive
Set was defined for each trial based on the instruction and the to-be-
memorized list, as shown in Figure 1. Once again, the latter information
came from a parameter tape which was read before each trial began. We
should also note that while the probability of a Yes response was constant,
the probability of displaying a particular stimulus was confounded with
set size (the larger the set size the less likely any one stimulus would
appear).

On those tests requiring a No response, one of three categories
of items could be selected. Once again a random number chart was used
by the computer to determine which of two pools of items should be
consulted. With probability 1/2 (overall p = 0.25), a test item was
selected from a pool of _Inclusive_ items, otherwise it came from the
larger pool consisting of both _Exclusive_ items and the 12 _Null_ Set
items. On those trials where there were no Inclusive items, the selec-
tion would automatically be from the second pool (Exclusive plus Null
items).

The term Inclusive is appropriate for two conditions in Experi-
ment I -- Looks and Means. An Inclusive item is defined as a homonym
of a displayed stimulus when that homonym version requires a No response
on the trial in question. If, as in Figure 1, the stimuli were the items
"WON, 2," the Inclusive Set would contain the items "1, ONE, TWO, TOO"
for the _Looks_ Condition, but only "1, ONE, TOO" for the Means, since
"TWO" means the same as "2" and consequently would be a member of the
Positive Set (requiring a Yes response).

Exclusive items are those remaining from the 12 potential stimuli
after all Positive Set items (correct for the trial in question) and
their homonyms (Inclusive items) have been removed. In the illustrated
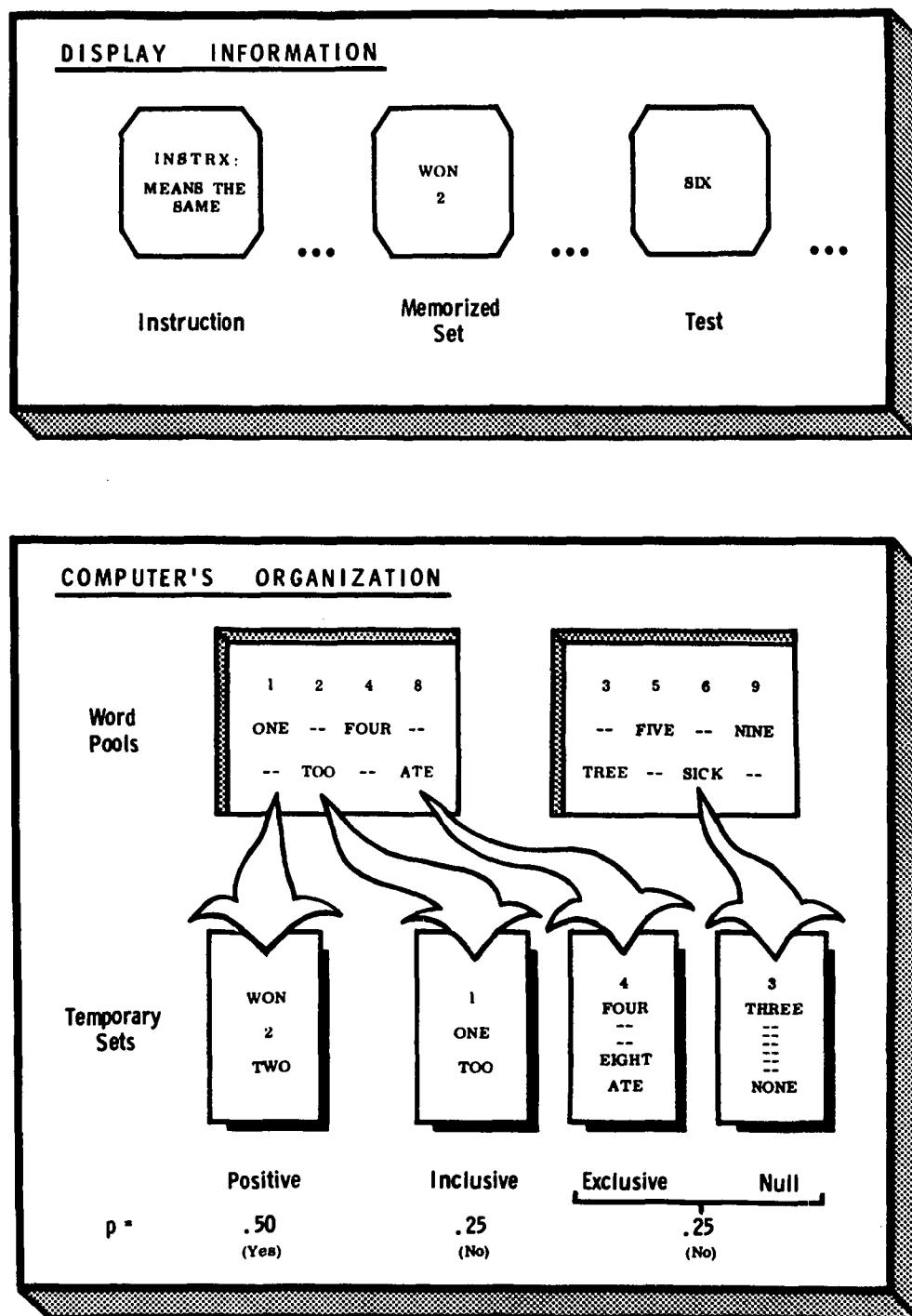
Fig. 1. Illustration of a trial, showing displayed information and the computer's arrangement of permanent word pools into temporary functional sets for the selection of test items.

example, this would include all members of the homonym triplets for the
sounds "4" and "8". Thus the Positive, Inclusive, and Exclusive Sets
must be redefined for each trial, being jointly determined by the instruc-
tions and displayed stimuli, whereas the Null Set is constant throughout
the experiment.

During a testing session, a subject responded for seven or eight
trials as the computer stored data. Then a rest period for the subject
was required while the data were punched onto tape. Next a summarized
version of the data was printed out via the teletype console. These
rest periods averaged nearly one minute each. Since a testing session
involved either seven (Experiments II and III) or eight (Experiment I)
practice trials, always followed by 48 data trials, the subject would
have six or seven rest periods per session. The session lasted 40-50
minutes.

PROCEDURE FOR EXPERIMENT I

The three conditions in Experiment I were the instructions
"LOOKS THE SAME" (Looks), "MEANS THE SAME" (Means), and "SOUNDS THE
SAME" (Sounds). The potential stimuli were as follows: 1, 2, 4, 8,
ONE, TWO, FOUR, EIGHT, WON, TOO, FOR, ATE. The Null Set was composed
of the items: 3, 5, 6, 9, THREE, FIVE, SIX, NINE, TREE, FINE, SICK,
NONE.

The interpretation of the Looks instruction is straightforward
in that the subject need only memorize the exact items which appear in
the to-be-memorized list. Test items identical to any one of the memor-
ized stimuli receive a Yes response. Consequently the Looks Condition
involves the same identification of Positive Set items as required by
the instructions used by Sternberg (1966) and those used by Wattenbarger
(1970) in his Physical Identity Condition.

The explanation of the subject's task in the Sounds and Means
Conditions will benefit from a digression to identify three variables
which characterize the stimuli. A stimulus set can be specified in
terms of (1) the number of items displayed in the to-be-memorized set = D,

(2) the number of potentially correct items = C, and (3) the number of "Types" represented in the display = T. The term "Types" refers to the categories: Arabic numbers, English spelled numbers, and non-number English words. Variables D and T are determined by the to-be-memorized list as it is displayed, and consequently these variables are not affected by the instruction. The variable C is determined by the instruction as it relates to the to-be-memorized list of stimuli. Table 1 illustrates these three variables in conjunction with the Sounds and Means instructions. For the Looks instruction, C = D. For all three instructions, D varies from 1-4 while T varies from 1-3. C takes the values 1-4 for the Looks Condition, 3, 6, 9, and 12 for the Sounds Condition, and 1-8 for the Means Condition.

In part, the instruction serves to define the D:C ratio. Since for the Looks instruction only the displayed items are correct, the D:C ratio is constant at 1:1. For the Means instruction the ratio is 1:1 only when the non-number words are displayed (see first two examples of Table 1). For the numerical stimuli, there is a 1:2 ratio between D and C (unless both forms of the number are displayed as stimuli, as in the third and fourth examples in Table 1). Thus displaying either numerical form, as in the first example of Table 1, signifies that both forms are correct, since they have equivalent meanings. Similarly, for the Sounds Condition, all members of a homonym triplet are correct when any one is displayed, resulting in a maximal 1:3 ratio (as in the first, second, and last examples of Table 1).

The trial sequence for Experiment I, which contains only a sample of all possible stimulus displays, was identical for all three instructions. [The contents of the trials are listed in Appendix A by trial number, whereas the sequence of trials is preserved in Appendices B, C, and D.] The sequence was arranged to maximize critical comparisons between the instructional conditions. The 48 data-collection trials were first divided equally among the four values of variable D (i.e., 12 trials displayed one item, 12 more displayed two items, etc.). For each value of D, the 12 trials were evenly divided among the possible values of C as defined by the Sounds instruction. For example, of the 12 trials where

TABLE 1

EXAMPLES OF VARIABLES D, T, AND C FOR

THE MEANS AND SOUNDS CONDITIONS

D	T	Display Example	Sounds Condition Positive Set	C	Means Condition Positive Set	C
1	1	ONE	1,ONE,WON	3	1, ONE	2
1	1	WON	1,ONE,WON	3	WON	1
3	3	4,FOUR,FOR	4,FOUR,FOR	3	4,FOUR,FOR	3
3	3	WON,8,EIGHT	1,ONE,WON 8,EIGHT,ATE	6	WON 8,EIGHT	3
3	1	1,2,8	1,ONE,WON 2,TWO,TOO 8,EIGHT,ATE	9	1,ONE 2,TWO 8,EIGHT	6

three items were displayed, four displayed all three members of a homo-
nym triplet such that C = 3, four displayed two sound classes such that
C = 6, and the remaining four trials displayed three different sounds
such that C = 9 (third, fourth, and fifth examples, respectively, from
Table 1). Finally, within a given D-C pair with C calculated for the
Sounds instruction, the Types of items were varied to include different
values of C as calculated for the Means Condition. For example, with
three displayed items incorporating two sounds, C values of three
(WON, 8, EIGHT), four (TWO, TOO, FOR), and five (ONE, WON, FOUR) were
represented in the Means Condition.

A difficulty with the above priority system is that while it
equalizes the number of responses for each value of D (and therefore of
C) within the Looks Condition, it creates an imbalance for the other
two conditions, having more cases with small values of C and very few
with C at its maximum. [This problem was adjusted in Experiment II by
counterbalancing first on the variable C, and then on D and T, thereby
equalizing the number of responses among the sets defined in auditory
units.]

To the limit of the total number of displayed items, all stimuli
occurred equally often in the to-be-memorized display. The order of
trials was determined from a random number table with the constraint
that no three consecutive trials contain the same displayed stimulus.
[A counterbalancing method of trial ordering was used in Experiment II
to distribute more evenly trials throughout a session.]

CHAPTER III

THREE CRITERIA

The results which best relate to our original question are found
in Experiment I. It is here that the three competing hypotheses will
be compared using identical materials but different instructions. Each
instruction was designed to be most compatible with a single hypothesis.
Presumably the Looks Condition calls to the subject's attention the
visual features of the stimuli. Similarly, the Means Condition forces
the subject to use the semantic features of the stimuli while the Sounds
Condition requires attention to the auditory properties of memorized items.

Our first interest is in the degree to which the different instruc-
tions facilitated the speed of responding. But the average latency of
responding does not provide a strong test of our hypotheses. The use of
the Sternberg paradigm allows us to compare slopes between conditions.
When reaction times are grouped according to the size of the Theoretical
Positive Set, we can derive the average time to make a single memory
comparison (i.e., the slope). Our three hypotheses suggest the form of
memory used for this memory-matching process. By making the naive assump-
tion that subjects search a visual memory in the Looks Condition, a seman-
tic memory in the Means Condition, and an auditory memory in the Sounds
Condition, we can interpret the slope of the Looks Condition as the esti-
mate of the rate of search through a visual store, the slope of the Means
Condition as the rate of searching a semantic store, and the Sounds Condi-
tion slope as reflecting the comparison time with auditory memory codes.
In each case these slopes are calculated using the Theoretical Positive
Set which is compatible with the instruction (i.e., Looks = Visual Hypothe-
sis; Means = Semantic Hypothesis; Sounds = Auditory Hypothesis).

While it is conceivable that there are three distinct forms of
memory, such that all three hypotheses are verified, the Law of Parsimony
recommends we ask: "What pattern of results would indicate the presence

of a <u>single</u> form of memory?" We can deduce that a single memory code is being used if we can meet the following three criteria.

I. One condition results in a shallower slope than the other two and the reaction times are at least as fast.

II. There is less variability among trials within each Positive Set size in the favored condition of Criterion I relative to the intra-set va_iability of the other two conditions.

III. The hypothesis compatible with the favored condition (by the above two criteria) is capable of reorganizing the trials of the non-favored conditions according to new Theoretical Positive Sets such that the resulting data meet two standards: (1) the newly calculated slope is equal to th_t of the favored condition, and (2) the variability among trials within a set has been reduced relative to the original calculation of intra-set variability. A further discussion of these criteria follows.

Although there is an instructional condition which is most compatible with each hypothesis, we can also apply the other "less-compatible" hypotheses to each condition. For each instructional condition we can order trials according to the Visual, Semantic, or Auditory Theoretical Positive Sets. One way to conceptualize this 3 x 3 set of possibilities is by an analogy to a deck of playing cards. Let us say that we have been dealt a hand of seven cards. Without substituting any cards we order this hand by first arranging them by number value, then by suit. Alternatively, we could have first arranged the cards by suit, and then within each suit by number value. The same cards can be ordered two ways depending on the game we wish to play, or on the highest point value within a game. In a similar way, we have a set of trials which can be organized by each of the three hypotheses (redefining the Theoretical Positive Set for calculating the slope) separately for each of the three instructions.

We can get a feel for the validity of the first criterion through a brief example. First let us assume that there is an ideal observer who has a true intercept of 400 msec. (the sum time for all processing in the absence of a memory comparison process). Our observer takes 50 msec. to make a single memory comparison. Thus when he has one stimulus code in memory his response time to a test is 400 + 50 = 450 msec. For set

size two, his time is 400 + 2(50) = 500 msec., 550 msec. for set size three, etc. Now assume that memory is visual. When our ideal observer works with the Looks instruction we calculate (via Visual Positive Sets) an intercept of 400 msec. and a slope of 50 msec. per item. The Sounds instruction demands that any homonym form of a to-be-memorized item also receive a Yes response (as in Table 1). With three versions for each homonym our subject has to store three items for each separate sound in the to-be-memorized list. Consequently, when one sound is correct the subject will have to search three visually-distinct memory codes resulting in a response latency of 400 + 3(50) = 550 msec. Two sounds requires six memory traces, yielding a 700-msec. response time, etc.

We can now illustrate the validity of the first and third criteria. Our first step is to plot the data of each condition according to its compatible hypothesis. This implies the Sounds Condition data would be plotted in auditory units, one unit for each memorized sound. The above time for our ideal subject was 550 msec. for one auditory unit, 700 msec. for two units, 850 msec. for three, etc. These latencies imply a slope of 150 msec. per sound unit, and an intercept of 400 msec. While the intercept is unchanged relative to the Looks Condition, the slope for the Sounds Condition is steeper than that of the Looks Condition when the latter was based on the Visual Hypothesis. Thus we have verified . Criterion I -- the shallowest slope occurs in the condition compatible with the correct hypothesis. Furthermore, if we reorganized the Sounds Condition data via the Visual Hypothesis, we can verify Criterion III. In this case we merely adjust the abscissa such that instead of one sound unit, it reads three visual units (since there were three versions of each sound). Two sound units incorporates six visual units, three become nine visual units, etc. The newly calculated slope is obtained by merely dividing the 150-msec. steps into three memory codes. The Sounds Condition slope becomes 50 msec. per visually distinct item according to the Visual Hypothesis. This value matches the result from the Looks Condition, confirming Criterion III.

Finally, let us consider an illustration of the second criterion, less variability. Suppose that in the Means Condition there are three trials -- X, Y, and Z. All three contain three semantic units. Trials X

and Y incorporate two auditory units while trial Z includes three auditory units. Let us further assume that memory is auditory. Consequently the response times for X and Y are nearly identical, but the time for Z is longer (by an amount equal to the time to search its one additional auditory unit). Grouping these three trials together under the Semantic Hypothesis creates unnecessary variability among trials of set size three. The range of response times is wide. However, when we reorganize the data of the Means Condition according to the Auditory Hypothesis, we correctly classify trials X and Y together as involving two sound units, while trial Z will be grouped with any other trials involving three sound units. The new Theoretical Positive Sets, defined in auditory units, show little variability among response times of similarly grouped trials. Moreover, we can calculate the slope of a line through the mean value for each set size. As in the previous example, we expect this slope to be equal to the slope of the Sounds Condition (calculated in auditory units).

Now that the appropriate means of evaluating the data has been outlined, we are prepared to launch into the results. First we shall consider the reaction time findings for Experiment I, to be followed by an analysis of errors.

EXPERIMENT I

Criterion I

The strategy for analysing the results of Experiment I will be to apply the three criteria to the data in order to identify a single form of memory, if such a conclusion is warranted. Only if this strategy fails will we conclude that more than one form of memory code is functional. As a preliminary step, the average weighted Yes and No response time was calculated for each trial, pooling the reaction times of the six subjects. Appendix A lists the items displayed on each trial while Appendices B, C, and D list unweighted trial means by series for the Looks, Means, and Sounds Conditions, respectively. [The weighted means preserve the information of every response whereas the unweighted means retain the average

time for each subject. In general, the weighted means will yield faster
reaction times.] A highly summarized version of the data is presented
in Table 2, where Yes and No response tiiu.2s have been pooled. These data
permit application of Criteria I and II. Considering the absolute magni-
tude of the reaction times, the overall mean times (based on 8,640 timed
observations) were 461, 494, and 513 msec. for the Sounds, Looks, and Means
Conditions respectively. An analysis of variance test verified that the
Sounds Condition was significantly faster than the other two conditions
(p < .01 between Sounds and Looks). On the basis of the overall mean times,
the criterion of lowest response times favors the Auditory Hypothesis.

Inspection of response times as a function of the size of the
Theoretical Positive Set (Table 2) reveals that none of the conditions
produced a linear slope through all four sets. However, the slope for
the Looks and Sounds Conditions are nearly linear through three of the
four points. This finding is more apparent in Figure 2 which displays
results of these two conditions by series. The slope for the Sounds
Condition (right side of figure) is clearly flatter, averaging approximately
30 msec./sound for sets 1-3 (the line accounting for 99.8% of the variance).
The drop in latency for set size four is perplexing. Also confusing is the
finding that the reaction time to a Theoretical Positive Set of one unit
is faster for the Looks Condition than for the Sounds Condition. If
memory is solely auditory, the latter should be at least as fast. We
shall return later to these two puzzling results.

The slope of the Looks Condition (left side of Figure 2) attenuates
from 54 msec./item to 45 msec./item with successive series. Unlike the
value in Table 2, the plotted point for set size three does not include
data from the four trials where all three versions of the same sound were
displayed (i.e., a complete homonym triplet). These data will be discussed
later. The curve for Series One is a near-perfect straight line when the
abscissa is plotted logarithmically (similar to the data of Swanson & Briggs,
1969). However, the curve flattens to a straight line by Series Three.
Series One's deviation from a linear function is related to the Yes response
to a single item, where the observed latency is 60 msec. faster than that
predicted from a straight line passed through the Positive Set points 2, 3,

TABLE 2

EXPERIMENT I REACTION TIME MEANS, STANDARD DEVIATIONS, AND RANGES AS A
FUNCTION OF SIZE OF THE THEORETICAL POSITIVE SET, LISTED BY CONDITION

Condition	Hypothesis	\Calculation	Theoretical Positive Set			
			1	2	3	4
Looks	Visual	Mean	413*	478	526	562
		S. D.	18	16	36	53
		Range	61	57	115	179
		# Trials	(12)	(12)	(12)	(12)
Means	Semantic	Mean	462	495	566	568
		S. D.	16	37	40	88
		Range	59	125	133	286
		# Trials	(15)	(14)	(10)	(9)
Sounds	Auditory	Mean	437	468	495	480
		S. D.	14	11	18	12
		Range	53	33	54	31
		# Trials	(22)	(14)	(8)	(4)

* Time in milliseconds
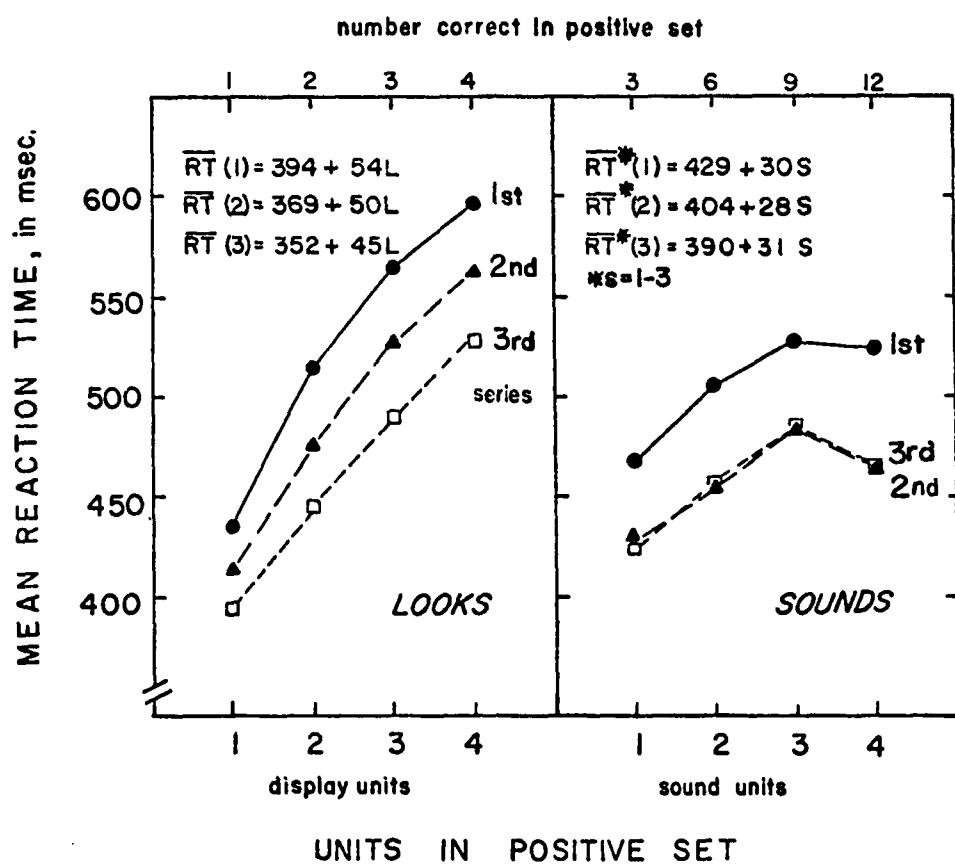
Fig. 2.   A comparison of reaction times as a function of units in the
Theoretical Positive Set for Looks and Sounds Conditions of
Experiment I, displayed by series. Slopes and intercepts for
linear regressions, recorded on the figure, were calculated
using theoretical units and the equation y = a + bx.

and 4, and extrapolated to the one-item case. Moreover, the slope of
this new line for Series One is 45 msec./item, which is identical to the
slope in Series Three. [Wattenbarger also found the Yes response in the
one-item case to be faster than expected -- the discrepancy being approxi-
mately 90 msec.] In the present study, the No latency was also faster
for the one-item trials in Series One, but only 20 msec. below the expected
time. This difference also disappears by Series Three. In conclusion,
practice does not appear to affect the slope, which we may take to be
about 45 msec. per visual item, but practice does affect the intercept
and the one-item response time.

In summary, the faster overall reaction time and the shallower slope
of the Sounds Condition, relative to the other two conditions of Experiment
I, supports the Auditory Hypothesis.

Criterion II

The second criterion suggests the inter-trial, intra-set variabil-
ity within the Sounds Condition (defining Positive Sets via the Auditory
Hypothesis) should be less than the variability within the other two
conditions if memory is solely auditory. Table 2 reports two measures
of variability: the intra-set variability or standard deviation of trial
means within each value of the Theoretical Positive Set, and the intra-set
range or difference in milliseconds between the fastest and slowest trial
means. On both measures the Sounds Condition is superior, further support-
ing the Auditory Hypothesis.

Criterion III

Ideally it would be preferable to reorder the data of each condition
according to the other two "less compatible" hypotheses, such that all
three hypotheses were compared within each condition. In practice, this
operation is difficult. Moreover, the slow reaction times, irregular
curve as a function of set size, and extreme variability within the Means
Condition when ordered by the Semantic Hypothesis eliminates the Semantic

Hypothesis from serious consideration. On the other hand, despite the steeper slope and greater variability of the Looks Condition (Visual Hypothesis) relative to the Sounds Condition, the fact that (1) the Visual Hypothesis produced a nearly linear slope in ordering the data of the Looks Condition, and (2) the absolute reaction time for the one-item trials was lower in this condition than in any other, encourages us to further examine the Visual Hypothesis. Therefore, Criterion III will be applied to both the Visual and Auditory Hypotheses. Since the process of reorganizing conditions via the Auditory Hypothesis is quite elaborate, and requires a refined version of this hypothesis (the need for which will hopefully become apparent after the discussion of later experiments), this analysis will be postponed until Chapter VI. We now consider the Visual Hypothesis as it applies to the Means and Sounds Conditions.

In the Means Condition the subject responds Yes to any item which has the same meaning as a memorized item. Therefore memorizing the item ONE implies a Yes response to either ONE or 1 (as illustrated in Table 1), since both have the same meaning. However, if WON were displayed, only WON would receive a Yes response. While each of the above instances includes one _semantic_ unit, the Visual Hypothesis states the former case involves _two_ distinct _visual_ units compared to one unit for WON. On this basis the trials of the Means Condition were reordered (according to the C variable -- the number of potentially correct, visually distinct items). The new Visual Positive Sets ranged from 1-8 items. The mean reaction times for each set, in ascending order, were 444, 477, 492, 557, 571, 562, 601, and 527 msec. A least squares line through all points results in a slope of 17 msec./item, but accounts for only 60% of the variance. Eliminating the last point yields a 26-msec. slope accounting for 92% of the variance. While this value appears impressively low, we must remember that the critical question is does it match the slope for the same hypothesis when applied to its compatible condition. The above value is only about half as great as the slope in the Looks Condition, failing to verify Criterion III.

A second aspect of Criterion III requires that the new organization reduce the intra-set variability. This requirement is difficult to judge,

since the Means Condition trials were divided among four sets under the
Semantic Hypothesis, relative to dividing the same number of trials among
eight sets with the Visual Hypothesis. In general, the more sets, the
fewer cases within each set and the lower the intra-set variability. The
extreme case -- having a one-trial set, hence zero variance -- does exist
in the present reorganization. Nevertheless, by noting that the newly-
formed set size six has the same range as the maximum range under the
Semantic Hypothesis (286-msec. difference between fastest and slowest
trials), we can conclude that the Visual Hypothesis has not contributed
to the predictability of the Means Condition data.

In applying the Visual Hypothesis to the Sounds Condition, it will
be instructive to review Figure 2. The abscissa for the Sounds Condition
(right side of figure) is plotted in auditory units on the bottom axis,
but in visual units (number of potentially correct items) at the top of
the figure. Since there are always three versions of each sound, the ratio
of visual to auditory units is always 3:1. This eliminates the need to
regroup trials. All trials remain in the same sets; we merely count each
of the four sets as having three times as many memory codes (i.e., 3, 6,
9, and 12 visual units). Consequently, the slope is effectively 1/3 the
value listed on the figure (previously calculated in auditory units), or
approximately 10 msec./item. Once again, while this visual slope is very
shallow, it does not match the slope of the Looks Condition, the latter
being 4.5 times as steep. In summary, the Visual Hypothesis fails to meet
Criterion III when applied to both the Means and Sounds Conditions. There-
fore, it is safe to conclude that memory is not solely visual.

Two Codes?

Although we can eliminate both the Semantic and Visual Hypotheses
from consideration as the sole form of memory codes, it is not yet possi-
ble to conclude that subjects fail to use a visual form of memory during
the Looks Condition. The evidence for this position comes from the two
previously cited findings: (1) the one-item trials for the Looks Condition
are faster than the one-sound trials in the Sounds Condition, and (2) the

slope of the Looks Condition is linear. Let us pursue this argument in greater detail.

If memory were solely auditory, then it should require the same time to search a one-sound memory trace regardless of the instructional condition. However, in the Looks Condition when a single item is correct, the subject must first store the sound of that item, and also a code for which version of that sound is correct, since there are three homonym forms for each sound. We shall refer to this additional code as a "tag." It is further assumed that it requires time to check this tag, relative to those instances where any version of the sound could be correct. Consequently, it should take slightly longer to search for a single item in the Looks Condition than a single sound in the Sounds Condition. Figure 3, which displays data from the Looks and Sounds Conditions separately for each subject, reveals that the reaction time for a single item in the Looks Condition was always faster than for one sound in the Sounds Condition. We should note that this finding has an ambiguous interpretation, since it could imply either (1) there is a visual memory which operates in the Looks Condition, although not in the other two conditions, or (2) the intercept of the Sounds Condition is higher than it would be if there were no homonym versions of each sound. The slope and intercept of the Sounds Condition in the absence of homonym stimuli will be evaluated in Experiment II.
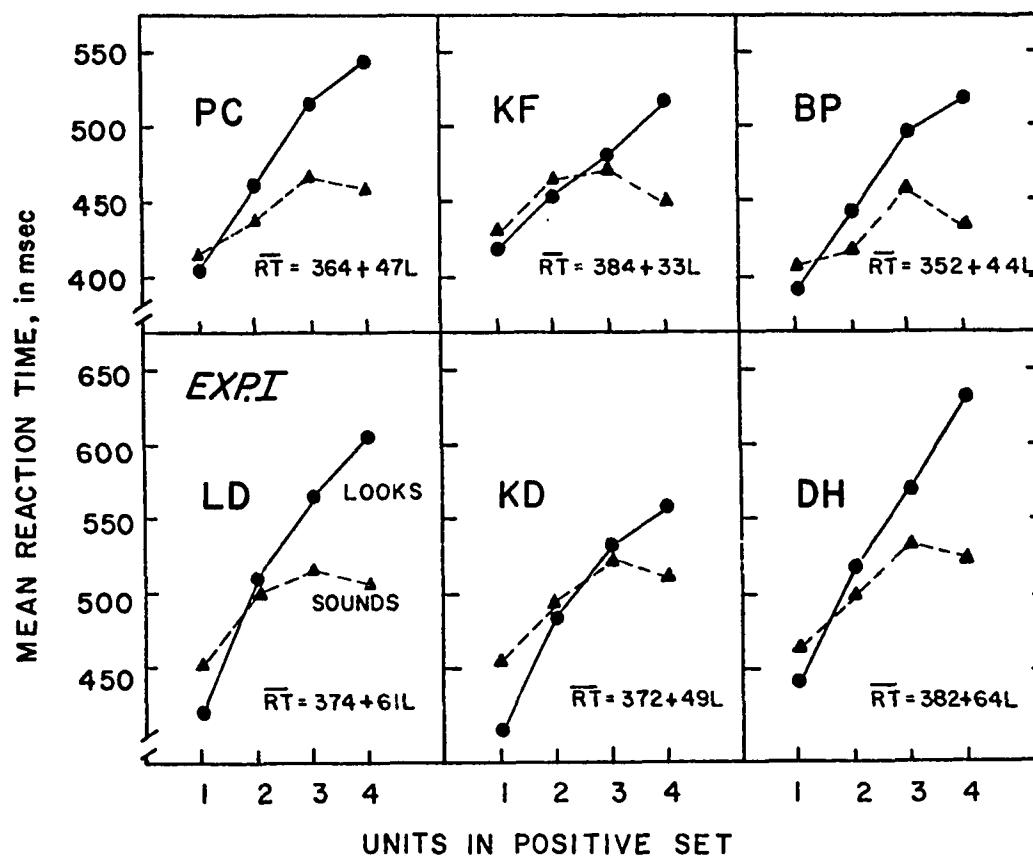
Fig. 3. Mean reaction times as a function of the size of the Theoretical Positive Set for the six subjects of Experiment I showing overall performance in the Looks and Sounds Conditions. The equations report the slope and intercept for a linear regression based on the Looks data.

ERRORS

Two questions concerning errors deserve our attention: (1) Do
the error rates differ among conditions within the same experiment?
(2) Does the pattern of confusions favor one or more of the competing
hypotheses?

Overall Rates

In answering the first question, we consult Table 3.   Although
our concern at the present time is only with Experiment I, for convenience
the overall errors for all conditions of all experiments are included.
In addition to recording total errors, Table 3 reports the proportion of
false positive to false negative errors.   A _false_ _positive_ is a Yes re-
sponse to a test item which should have been classified as No -- that is,
the item is _not_ a member of the Positive Set.   Similarly, a _false_ _negative_
is an incorrect No response to a member of the Positive Set.   The ratio of
incorrect Yes to No responses is reported in the right-most column of
Table 3.

In Experiment I, there is little discrepancy between conditions in
total errors.   The slightly higher rate for the Looks Condition (3.5% vs.
2.9% and 2.8% for the other two conditions) can be fully attributed to the
first session performance of a single subject who made 41 errors.   Therefore,
we conclude the overall error rates did not produce evidence favoring any
of the conditions of Experiment I.

Another important pattern of errors is the change in error rate as
a function of the size of the Positive Set.   Table 4 reports the proportion
of errors according to the size of the Positive Set for all but the Means
Condition.   As in the findings of Sternberg and Wattenbarger, error rates
increase as the number of to-be-memorized items increases.

TABLE 3

FALSE POSITIVE, FALSE NEGATIVE, AND TOTAL ERROR RATES

FOR ALL EXPERIMENTS

Experiment	Condition	Overall Rate	Number of Errors		Ratio (FP/FN)
			False Pos.	False Neg.	
I	Looks	3.5%	139	162	0.86
	Means	2.8%	130	115	1.13
	Sounds	2.9%	101	147	0.69
II	Sounds 1:3	3.5%	129	71	1.82
	Type 1:3	5.0%	172	116	1.48
	Sounds 1:1	5.2%	152	150	1.01
	Exten. 1:1	5.2%	149	153	0.97
III	Looks	3.9%	46	48	0.96

TABLE 4

PERCENTAGE OF ERRORS BY CONDITION AS A FUNCTION

OF THE SIZE OF THE POSITIVE SET

Condition	Experiment	Units in the Positive Set					
		1	2	3	4	5	6
Looks	I	2.0	3.8	3.5	5.1		
	III	1.3	2.0	2.0	3.3	4.5	10.5
Sounds 1:3	I	2.8	2.9	3.5	2.4		
	II	3.1	4.3	4.3	2.4		
Sounds 1:1	II	2.0	4.1	5.0	5.0	6.6	8.0
	Ext.	3.3	4.5	4.9	5.1	6.7	7.0
Type 1:3	II	3.8	6.9	6.0	3.2		

Types of Errors

Confusion matrices, similar to these discussed in Chapter I, were
constructed to evaluate our three competing hypotheses. This approach
assumes that errors associated with the sound of an item imply an auditory
memory code. To be sure, some errors may result from misidentification
while others may involve "precipitated motor actions." For simiplicity,
we will assume the errors were produced by difficulties in the memory
comparison process. For this analysis 166 false positives and 223 false
negatives from the Looks Condition plus 166 false positives and 147 false
negatives from the Means Condition are reported. All data come from
Experiment I and include the eight practice trials of each session. In
general, the 702 errors provide undeniable support for the Auditory Hypo-
thesis, especially the false positives which were overwhelmingly to homonyms.
Specifically, 288 false positives were to homonyms vs. 44 to non-homonyms.
The expected error ratio of homonyms to non-homonyms was 13:15. There is
little support for the Visual and Semantic Hypotheses. The remaining pages
of this chapter detail the various comparisons within confusion matrices re-
lating to the three hypotheses. The disinterested reader may wish to
skip this section and go on to Chapter IV.

The three hypotheses are evaluated in eight separate analyses. The
Auditory Hypothesis has four tests including both false positives and false
negatives separately within the Looks and Means Conditions. The three
evaluations of the Visual Hypothesis include false positives for both Looks
and Means Conditions and false negatives in only the Looks Condition.
Finally, the Semantic Hypothesis generates one test using a subset of false
positives only from the Looks Condition.

In nearly every analysis it is difficult to establish the expected
frequencies for the cells of the confusion matrix. Since the computer
generated the tests according to preestablished probabilities, the exact
number of opportunities for each type of error can only be estimated.
In some cases, complex conditional probabilities are needed, based on
the exact contents of each trial. In all cases, these expected values
were estimated by first assuming all possible types of trials were repre-

sented equally often, and then correcting these estimates in light of
the actual distribution of trials.

Auditory Hypothesis. The Auditory Hypothesis predicts more false
positive errors to incorrect homonyms of a memorized item than to incorrect
non-homonyms. As shown in Table 5, both the Looks and Means Conditions
produced a total of 166 false positives. Of this total, 142 for the Looks
Condition and 146 for the Means Condition were to homonyms, clearly
differing from the expected frequency of 77 errors (the smaller X^2 = 102.34,
p < .001) and supporting the Auditory Hypothesis. The expected frequency
was calculated in the following manner.

A false positive requires that the test item not be a member of
the Positive Set, that is, the test should receive a No response. As
discussed in Chapter II, and illustrated in Figure 1, there are three
types of items which can receive a No response: (1) a homonym of a Posi-
tive Set item (Inclusive Set), (2) a potential stimulus which is neither
a member of the Positive Set nor of the Inclusive Set (i.e., it is an
Exclusive Set item), and (3) an item from the Null Set. For present pur-
poses the last two categories will be grouped together since they are both
non-homonyms. As indicated in Figure 1, the probability of a test being
an incorrect homonym (Inclusive Set) was .25, compared to a probability
of .25 that the test item was a non-homonym. Homonyms requiring a No re-
sponse occurred on 52 of the 56 trials (four trials display all three
homonym versions of a sound). Therefore, the probability of a non-homonym
was .25 for 52 trials but .50 for the remaining four trials (when no
Inclusive items were present the selection of an item for a No response
automatically came from the Exclusive-Null pool). These constraints lead
to an unbiased expected ratio of (1/4)(52) = 13 homonyms for every
(1/4)(52) + (1/2)(4) = 15 non-homonyms. Since there are 166 false positives
in both the Looks and Means Conditions, the expected number of homonyms
vs. non-homonyms is 77 vs. 89.

The Auditory Hypothesis can also be tested by reference to the nature
of the stimulus set. False negatives are separated into the four trials
where all homonym versions were displayed (complete triplets) vs. those
(52) trials where some, but not all, versions were displayed (incomplete

TABLE 5

OBSERVED NUMBER OF ERRORS AND UNBIASED EXPECTANCIES
WITH REFERENCE TO AUDITORY CONFUSIONS

Error		Looks obser.	Looks expect.	Means obser.	Means expect.
	(TEST ITEM)				
False Positive	Homonym	142	(77)	146	(77)
	Non-homonym	24	(89)	20	(89)
	(STIMULUS SET)				
False Negative	Complete Triplet	7	(16)	11	(10)
	Incomp. Triplet	216	(207)	136	(137)

triplets). The Auditory Hypothesis predicts it should be easier to remember that all versions of a sound were present relative to keeping track of which visually-different representations were displayed and which were not. Consequently, there should be _fewer_ errors to complete triplets than to incomplete triplets. Table 5 reports a 7:216 distribution (complete:incomplete) for the Looks Condition and an 11:136 distribution for the Means Condition. Neither distribution departed from the unbiased expected number of false negatives, failing to support the Auditory Hypothesis. However, this comparison merely compared the four trials with triplets to all other trials. Twelve of these 52 trials contained only a single item. Table 4 has already revealed that these trials produced fewer errors. Consequently, a more appropriate comparison employs only those trials which have an equal number of memorized items. Excluding practice trials, eight trials displaying three items were incomplete triplets, four were complete. This 2:1 ratio was clearly exceeded in the obtained data from the Looks Condition where the false negative ratio was 39:7 ($X^2 = 6.66$, $p < .01$). The similar comparison for the Means Condition was not significant. In summary, the evidence from false positives provides unequivocal support for the Auditory Hypothesis while the evidence from false negatives provides mild support, coming only from the Looks Condition.

Visual Hypothesis. The Visual Hypothesis predicts false positives will occur more often to test items which begin with the same first letter as a memorized item, relative to items beginning with a different first letter. Since there were so few Exclusive and Null Set items involving false positives, only homonym errors were used to check for visual confusions. Table 6 reports 28 same-first-letter errors for the Looks Condition and 35 same-first-letter errors for the Means Condition. Neither value departs significantly from the expected number, failing to support the Visual Hypothesis.

The expected frequencies were calculated by determining the expected number of each type of test for each trial. A test item could either have the same first letter or a different first letter when compared to the member of the Positive Set for which it was a homonym. Such a comparison was

TABLE 6

OBSERVED NUMBER OF ERRORS AND UNBIASED EXPECTANCIES
WITH REFERENCE TO VISUAL CONFUSIONS

| Error | Condition | | First Letter of Test | | Arabic |
			Same	Different	Number
False	Looks	observed	28	59	55
Positive		expected	(33)	(65)	(44)
	Means	observed	35	85	26
		expected	(33)	(67)	(45)
False	Looks	observed	71	77	75
Negative		expected	(74)	(74)	(74)

only meaningful when both the test and the member of the Positive Set were words. If the test was a word, but the Positive Set homonym was an Arabic number, the test was classified "different." If the test item was an Arabic number (and a homonym), it was listed separately (right-most column). False positives to both words with a different first letter and to Arabic numbers (where the Positive Set homonym was necessarily a word) represent contradictions of the Visual Hypothesis, since neither could be visually similar to items requiring a Yes response.

The Visual Hypothesis was also tested for false negatives in the Looks Condition, as shown in Table 6. If words are encoded according to their visual pattern with special attention to the first letter, we expect an error in either (1) the recognition of a test word beginning with the same first letter as the stimulus (misperception of the test), or (2) the initial encoding of the stimulus, such that another word beginning with the same letter was stored in its place (misperception of the to-be-memorized word). Consequently, the Visual Hypothesis predicts more false negatives to words which have another stimulus involving the same first letter (as TWO-TOO, FOUR-FOR) relative to either unique-first-letter words (as ONE, WON, EIGHT, ATE) or to Arabic numbers. There is no bias among false negatives for any of these three equally-represented categories, further disconfirming the Visual Hypothesis.

Semantic Hypothesis. The Semantic Hypothesis was evaluated using the 55 false positives of the Looks Condition where the test item was an Arabic number. Since the Arabic homonym version should have received a No response, the Positive Set must have displayed (1) the non-number version of the sound, or (2) the spelled number version of the sound, or (3) both the non-number and spelled versions. If memory preserves the meaning of the stimulus, the incorrect response to the Arabic would occur more often when the stimulus set contained the spelled number than when it contained the non-number (since the former is a synonym). Table 7 reports the observed frequencies of each Type of stimulus set. The obtained values of 11, 21, and 23 errors depart significantly from expectancies based on the tally of 20, 13, and 15 trials containing non-numbers, spelled numbers, and both

TABLE 7

FREQUENCY OF INCORRECT YES RESPONSES TO ARABIC HOMONYMS BY TYPE

OF MEMORIZED ITEM IN THE LOOKS CONDITION, EXPERIMENT I

Series	Non-Number	Spelled Number	Both
One	6	8	13
Two	2	7	7
Three	3	6	3
Total	11	21	23

Types, respectively. While these data appear to support the Semantic
Hypothesis, a more detailed analysis reveals that 11 of the 21 errors
to spelled numbers come from only two trials. Trial 42 displayed the items
2, 4, and EIGHT (false positive response to test item "8") while Trial 58
displayed the items ATE, FOUR, 8, ONE, 2, and FOR (false positive response
on this practice trial to the test item "1"). In both cases there are
two other items in memory coded as Arabics. There were no matching trials
where the critical item (i.e., EIGHT for Trial 42 and ONE for Trial 58)
appeared in the non-number version. Thus the bulk of the evidence for
semantic confusions is heavily dependent on two trials which produced
unusual difficulty in keeping track of which version went with which
sound. This same evidence supports the alternative explanation that subjects
require additional tags to denote which version of a homonym is correct.

Additional evidence that subjects use tags specifying the Type of
homonym in question was provided by false positive responses to Null Set
Arabics. Four of the seven errors were on Trial 44, where the Positive
Set was composed of the items 1, 2, 4, and 8. The average reaction time
for the errors was 377 msec., which was 105 msec. faster than the mean Yes
time for this trial (for Series One and Two which contained these errors).
Another error was on the similar (practice) Trial 57 (Positive Set = 1, 2,
4, 8, ONE, and TWO; response time = 374 msec.). The other two errors.
were also on trials containing Arabics (Trial 13 = 1, ONE; Trial 10 = 8).
Both involved precipitated responses (times of 382 and 272 msec., respec-
tively). While these errors can be interpreted as evidence for the Visual
Hypothesis, the lack of converging evidence, the small number of cases,
and the extremely brief response times hardly inspires confidence in this
interpretation.

CHAPTER IV

QUESTIONS ARISING FROM EXPERIMENT I

This chapter confronts several technical difficulties in interpeting the data from the principal experiment. Five questions provide the impetus for the design of Experiments II and III. In general, these additional experiments increase our confidence in the previous conclusion that memory is auditory.

The findings of Experiment I, for both response latencies and errors were interpreted to support the Auditory Hypothesis. Nevertheless, some of the findings were difficult to explain if memory were solely auditory. In particular, we ask: (1) Could the data which appears to support the Auditory Hypothesis have an alternative interpretation? (2) In the Sounds Condition, why are the response times to four sounds faster than to three sounds? (3) Would the slope of the Sounds Condition be the same if there were no homonym versions for each sound? (4) Why is the average reaction time to one item in the Looks Condition faster than the time for one sound in the Sounds Condition? (5) Was the steep slope of the Looks Condition (relative to the slope of the Sounds Condition) influenced by the presence of other instructions -- specifically, the emphasis on the auditory properties of the stimuli due to the use of homonyms?

The last three questions are directly resolved by developing a set of stimuli which include no homonyms and then presenting the materials with both the "sounds the same" instruction (Experiment II, Sounds 1:1 Condition) and the "looks the same" instruction (Experiment III). We shall consider these experiments presently.

The first two questions require that we inspect a reasonable alternative to the Auditory Hypothesis. In Experiment I, an analysis of a set of trials suggested such an alternative explanation. Four trials displayed all three versions of a homonym triplet (the trials were numbered 31, 32, 33, and 34, although they were scattered in the trial sequence). For all three instructions the subject's task was the same -- to respond Yes when

any one of the three members of the triplet appeared as a test item.
Consequently, we expect the average response latencies for these trials
to be the same across the three conditions of Experiment I. The obtained
mean pooled response times for the Looks, Means, and Sounds Conditions
were 466, 460, and 427 msec., respectively. These differences were largely
due to Yes response latencies, which produced times of 455, 441, and 404
msec., compared to No response times of 477, 478, and 451 msec. (same order
of conditions in each case). For all six subjects the Yes and No response
times were fastest in the Sounds Condition. Note that the rank ordering
of conditions parallels the number of "chunks" or "groupings" implied by
the instructions. The Sounds Condition implies the three stimulus items
having the same sound are chunked into a single unit, so each of the Trials
31-34 would represent one chunk. The Means Condition would take two chunks
to specify the three items (numbers and non-numbers); and the Looks Condition
would need three chunks. The fewer chunks, the faster the response latency.
Since chunking items is known to facilitate recall (Miller, 1956), perhaps
a part of the effect we have observed in the Sounds Condition was not due
to the auditory properties of the items at all. Maybe it was merely the
result of a learned habit of chunking classes of items. We now entertain
an alternative model to the Auditory Hypothesis which may account for the
same findings. The "Chunking" Model proposes that the greater efficiency
for responding with large numbers of potentially correct items in the
Sounds Condition, relative to the Looks Condition, is due to the application
of a chunking rule, not an auditory memory code.

The Chunking Model does not take a position concerning the specific
nature of a memory code with reference to visual, semantic, or auditory
properties. It merely assumes that units of memory can be clustered to-
gether via a new (abstract?) code. Consequently, we do not have a fourth
hypothesis, as such. The possibility of chunking is an independent consider-
ation which can serve as an extension of any of the three major hypotheses.
The present concern is that a chunking strategy might have contributed to
the flat slope of the Sounds Condition, in effect undermining the rationale
for Criterion I. The obtained slopes eliminate both Visual + Chunking and
Semantic + Chunking possibilities, but leave three alternatives: (1) purely

Auditory codes, (2) Auditory + Chunking, (3) solely Chunking of abstract codes. Experiment II was designed to differentiate these alternative explanation..

How might chunking be effective? We might assume that subjects can group items into semantic equivalents and into auditory equivalents. The stimuli of Experiment I contain eight "meaning-chunks" or four "sound-chunks." Furthermore, we assume the subject can discriminate those items which are potential stimuli from those of the Null Set. For example, the Chunking Model interprets the fast reaction times on the four-sound trials in the Sounds Condition as the result of the more general two-set discrimination. The Chunking Model proposes that the subject learns to discriminate those items which are possibly correct stimuli from those of the Null Set. The four-sound trials are unique in that any item which is ever correctly classified as Yes will require a Yes response during these trials. Consequently, the subject need only search through one memory chunk on the four-sound trials. Consistent with this interpretation is the finding that the drop for the four-sound trials, relative to the three-sound trials, increases with greater familiarity for the stimulus items (i.e., across series). Figure 3 verifies that all six subjects were faster to four than to three sounds. However, no subject was as fast on four-sound trials as one-sound trials. Since both sets are presumed to involve a single chunk, the slower time to four-sound trials is inconsistent with an extreme version of the Chunking Model.

Another difficulty with the Chunking Model arises in the interpretation of the linear slope of the Sounds Condition for Positive Set sizes 1-3. The argument associated with sets 1-4 goes as follows. Suppose the subject has five chunks or categories: one for each of the four homonym triplets, and a fifth for the Null Set. Then, on those trials with a Theoretical Positive Set of two units, it would be most efficient simply to recall the two correct sounds. However, on the trials where three sounds are memorized, it would be most efficient to store just two chunks -- the homonym triplet which is incorrect, and the Null Set. Thus both types of trials would require memorization of two sets or two chunks. This reasoning leads to an expected function shaped like an inverted "U" -- equally fast times for set sizes 1 and 4, and equally slow times for set sizes 2 and 3. The data of Experiment I indicate that this was apparently not done.

The data from a new condition, designed to promote chunking according to some rule other than auditory similarity, could take one of three forms: (1) identical to the Sounds Condition results, which would invalidate the 'pure' auditory code interpretation, (2) similar to the expected results of the Chunking Model (inverted "U"), but unlike the Sounds Condition, which would establish that chunking can take place, that chunking could have produced the faster latencies on four-sound trials due to a two-set discrimination, and that the previous interpretation of auditory memory codes remains unchallenged by the Chunking Model, and (3) altogether different from the two expected results, which defies interpretation. The Type 1:3 Condition of Experiment II was designed to evaluate these possibilities.

EXPERIMENT II

Conditions

The three conditions of Experiment II included: (1) Sounds 1:3, a replication of the Sounds Condition of Experiment I, (2) Sounds 1:1,.a "sounds the same" instruction applied to new stimulus materials devoid of homonyms, and (3) Type 1:3, a rearrangement of the stimulus materials of the original Sounds Condition designed to evaluate the Chunking Model. Each subject was exposed to each condition only once.

The Sounds 1:3 Condition (the 1:3 representing the maximal D:C ratio) was included to serve as a control-baseline for interpreting the other two conditions of this experiment. It differed from the earlier trial sequence in that the counterbalancing equated the number of trials for each value of stimulus variable C, rather than variable D, as used in Experiment I in order to favor comparisons within the Looks Condition. Consequently, there were an equal number of trials representing 1, 2, 3, and 4 sounds. Up to six items were displayed, permitting D-C pairs of 1-3, 2-3, 3-3; 2-6, 4-6, 6-6; 3-9, 4-9, 6-9; 4-12, 5-12, and 6-12.

The Sounds 1:1 Condition used words which did not have homonyms plus the non-numerical items from Experiment I. As a result, there was one item correct for each item displayed (thus the D:C ratio was 1:1). The size of the Positive Set was expanded to six items, yet with 12 potential stimuli, there were always residual members of the set of potential stimuli which could serve as No response items on a given trial. This eliminated the possibility of a two-set discrimination effect. All items were three-letter words, which avoided discrimination based on word-length. [The mean reaction time to Arabic numbers in the one-item Looks Condition in Experiment I was 21 msec. faster than the time for either Type of spelled word.] The set of potential stimuli which contributed to the Positive Set for the Sounds 1:1 Condition included: WON, TOO, FOR, ATE, CAT, DOG, COW, PIG, ARM, LEG, EAR, and TOE. The Null Set items were: BOB, JIM, TOM, JOE, FLY, ANT, BEE, BUG, OAK, ELM, ASH, and PEA.

The Type 1:3 Condition reordered the original stimulus matrix of four sounds each with three Types, into a matrix involving four Types each with three sounds. Three Types were the same as in Experiment I -- Arabic numbers, English spelled numbers, and English non-number words -- to which Roman numerals were added. The pool of potential stimuli included the items: 1, 2, 4; I, II, IV; ONE, TWO, FOUR: WON, TOO, FOR. Since the subject's task was to respond Yes to any Arabic number if an Arabic appeared in the to-be-memorized display, the Null Set could not contain any Arabic numbers. The intention was to make the stimulus and Null Sets as distinct as possible, and to provide labels to the subjects to help define the sets. This should have allowed the Type 1:3 Condition every advantage in being as efficient as the Sounds 1:3 Condition, according to the Chunking Model. If the results show faster reaction times for the Sounds Condition, we will conclude that the Auditory Hypothesis better describes the previous findings. To maximize discriminability, Null Set items included: &, %, $, +, =, <, MKP, NHM, LFD, LUI, VIN, MOI.

Subjects were given the same instructions for the Sounds 1:3 and 1:1 Conditions. They were told that previous research showed subjects do best if they pronounce aloud the different sounds in the to-be-memorized list, and that they should rehearse the items silently. The possible Positive

Set sounds were described verbally to each subject before the testing
commenced. The instructions for Type 1:3 included an illustration in-
forming the subject that there were four sets of items which had to be
remembered, each set containing three members. Since the instruction on
the display scope for this condition read, "THE SAME SET," the subject
had to respond Yes to any item within a memorized set. The labels
"arabic," "roman," "english #," and "english words" appeared beside their
respective sets. [Appendix F contains the illustrated instruction sheet.]
These labels were recommended to the subject, with the further comment
that the subject was free to use any labels he chose, "such as A, B, C,
or D -- as long as they work." A sample trial was typed out, and the
subject studied the sheet for three minutes before beginning the seven
practice trials. Subjects were further instructed to call aloud the sets
which were present while the to-be-memorized display was on the scope,
and then to rehearse the labels. All subjects used the suggested labels
except one who referred to the spelled numbers as "American numbers."

Results

There was considerable variability among subjects within two con-
ditions of Experiment II, creating difficulties of interpretation. However,
an analysis of sequential effects (Chapter V) revealed that the data were
consistent for non-repetition responses, but varied in an unpredictable
manner for response repetitions. In addition to being consistent across
subjects and across testing sessions, the latencies of the non-repetition
responses agreed with expected values for both slopes and intercepts.
Consequently, we will not dwell in the present chapter upon puzzling details
which will be resolved in the next chapter.

Our first concern is whether the data from the naive subjects are
comparable to those of the trained subjects in Experiment I. The Sounds 1:3
Condition provides this test. When the results of Figure 4 are compared
with those of Figure 2, we verify that the average reaction time was nearly
identical for equivalent practice. The mean reaction time in Experiment I
for Series One, disregarding set size, was 495 msec., compared to 499 msec.
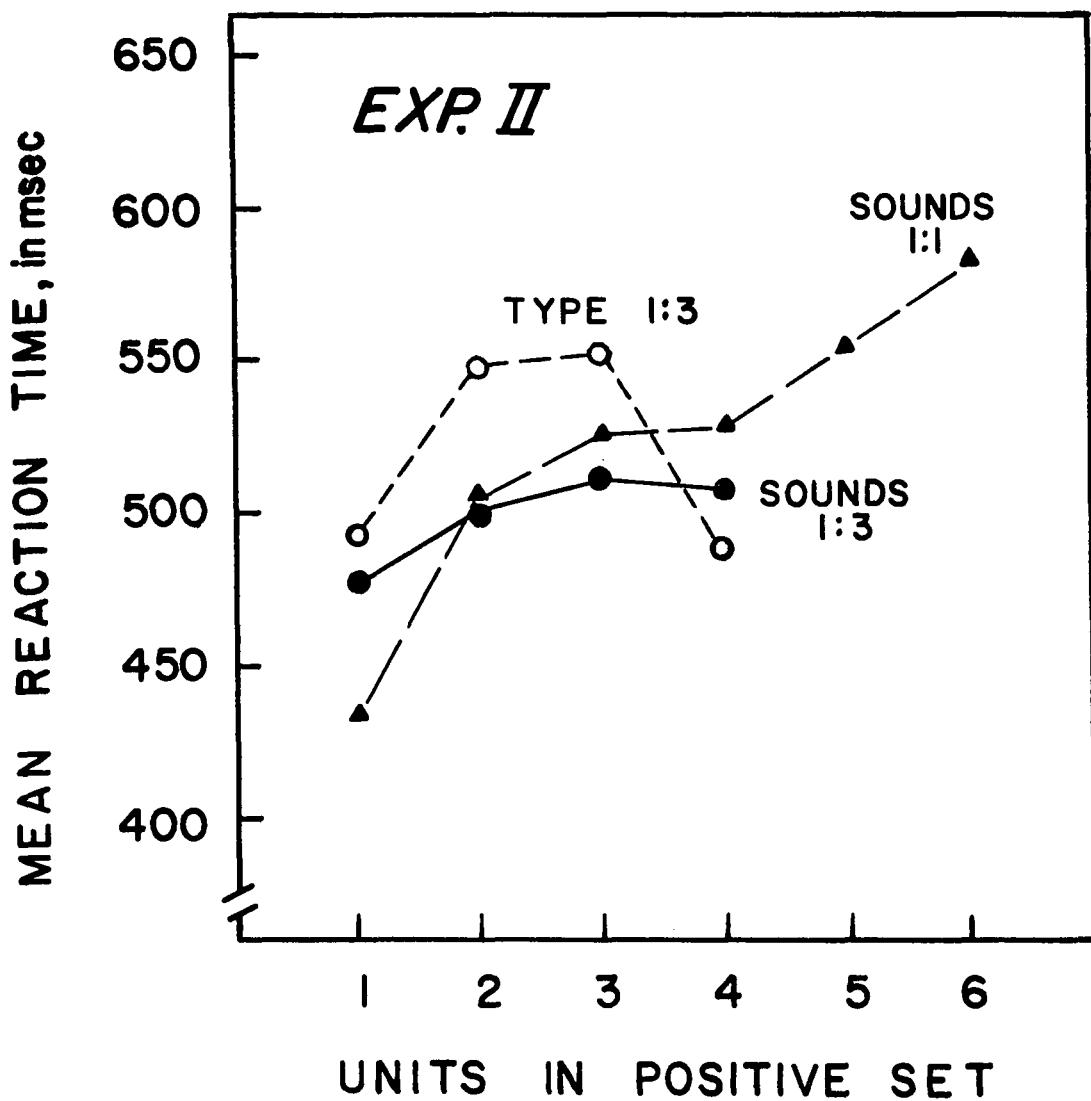
Fig. 4. A comparison of all conditions of Experiment II, plotting response times as a function of the Theoretical Positive Set. The data represent 13 subjects per condition, irrespective of hour.

in Experiment II. The slope in Experiment II across sets 1-3, however, was 17 msec./sound, considerably below the 30-msec. value of Experiment I.

Over comparable trials for both experiments, the Yes response slope in Experiment II (30 msec. per sound) was equivalent to that of Experiment I, although the No response slope (8 msec. per sound) was dissimilar. The fact that the intercepts of Yes vs. No responses differed by 101 msec. in Experiment II (compared to 57 msec. in Experiment I) suggests Experiment II subjects may have concentrated more on making a fast Yes response, such that all No responses tended to be made at more nearly the same latency.

Figure 4 shows the obtained results fit well with the inverted "U" pattern predicted by the Chunking Model for the Type 1:3 Condition. The interpretation of these data is: (1) chunking can affect response latencies in a predictable manner, yet (2) chunking alone can not account for the results of the Sounds 1:3 Condition. While the reaction time to one unit in the Theoretical Positive Set was only 15 msec. slower for Type 1:3 relative to Sounds 1:3, the difference increased to 50 msec. for set sizes 2 and 3. The faster time for the Sounds Condition was obtained for 10 of the 12 subjects serving in both conditions. The two subjects with slower latencies for the Sounds Condition had served in this condition for their first hour. Thus the advantage of the Type Condition for these subjects may simply reflect the effects of practice.

The faster time for No responses with set size four for the Type Condition is consistent with the Chunking Model. We assume the subject's strategy for the four-Type trials involved memorization of the single Null Set chunk, which required the No response, rather than the memorization of the four chunks requiring a Yes response. Although the No response was still slower than the Yes, the difference was reduced from 60 msec. for set size one (memorize one Yes response chunk) to 13 msec. for set size four (memorize one No response chunk). Only the Yes-No difference for set size four failed to reach statistical significance. Also, the No response time for set size four was significantly faster than the No times to all three other sets. By comparison, in the Sounds 1:3 Condition the fastest No response times occurred for set size one.

In addition to faster response latencies, the Sounds 1:3 Condition was superior in accuracy. Incorrect responses were made to 3.5% of the tests for the Sounds Condition compared to 5.0% for the Type Condition.

The results of the Type 1:3 Condition relate to the first two questions posed at the outset of this chapter. First, could the results of the Sounds 1:3 Conditions of Experiments I and II merely reflect the effects of a chunking strategy? The pattern of results from the Type 1:3 Condition led to the conclusion that that condition reflected "pure" chunking. Since the Sounds 1:3 Condition was significantly different from the Type Condition, it must have involved some other kind of processing. We reject the possibility that the results of the Sounds Condition can be solely attributed to chunking. However, we should not overlook the similarity between the Sounds 1:3 and Type 1:3 Conditions. Specifically, both produced a drop in latency for sets containing four units, although this drop was significantly greater in the Type Condition. This similarity suggests that while the Sounds Condition did not solely reflect chunking, it may have been susceptible to a two-set discrimination effect due to the selection of stimulus materials which incorporated a fixed set of four sounds. Therefore, in answer to our second question concerning the drop in latency for the four-sound trials, we conclude this result is an artifact of the stimulus materials. Verification of this conclusion is provided by the absence of such an effect in the Sounds 1:1 Condition which did not permit such a two-set discrimination.

The results of the Sounds 1:1 Condition were somewhat irregular. Under the Auditory Hypothesis, we expect the slope of this condition to be identical to that of the Sounds 1:3 Condition (i.e., 17 msec./sound). Although a slope of 19 msec. per sound was obtained across set sizes 2-6 (26 msec. for sets 1-6), the data points do not adequately match those of the Sounds 1:3 Condition. Since there were fewer visual patterns per sound, the stimulus uncertainty was reduced (i.e., the probability that a given item from the Positive Set would appear as a test item increased). This effect could result in a shorter time interval needed by a recognition stage of processing, in which the test item is transformed into an appropriate code for the memory comparison process. As a result the total time for a response would be reduced, producing faster latencies in the Sounds 1:1

Condition. Yet Figure 4 shows that only set size one resulted in a faster time for the 1:1 Condition, and that point was discontinuous with the rest of the function. In fact, this point was 2 msec. faster than the time for the Looks Condition, Series One of Experiment I. This similarity to the Looks Condition suggests the subjects may have been functioning in the same way during both conditions on those trials displaying a single item.

Large between-hour differences are shown in Figure 5. The term "hours" refers to the order of participating in each of the three conditions of Experiment II. Unlike the between-series comparisons of Experiment I, between-hour comparisons are necessarily between-subject comparisons, since each subject served in each condition only once. For all three hours, those trials with a single to-be-memorized sound produced exceptionally fast reaction times. Moreover, the data from Hours Two and Three with Sounds 1:1 were consistent and parallel to the Sounds 1:3 Condition for the critical set sizes of two and three sounds. By contrast, the data for Hour One appears to be more like the results of the Looks Condition, although the obtained slope is less than 31 msec./sound (24 msec./sound for sets 2-6). One interpretation of these changes relies on the fact that all subjects serving Hour Three in the 1:1 Condition have already served in the 1:3 Condition. Presumably, the latter condition forced a greater reliance on the sound properties of the memorized items. This task may have aided these subjects in the use of auditory representations, relative to those subjects just beginning the experiment. In preview of Chapter V, we can anticipate that these between-hour differences are mainly varied reactions to response repetitions, with the more experienced subjects making faster responses to repetitions than the less experienced subjects.

Before relating the results of the Sounds 1:1 Condition to our five questions, it will be to our advantage to satisfy ourselves that these findings are reliable. The Extension 1:1 Condition fulfills this need.

Extension 1:1

In order to obtain a more stable set of data for the Sounds 1:1 Condition, three subjects from Experiment I each participated for four hours in the Sounds 1:1 Condition. These subjects received the same in-
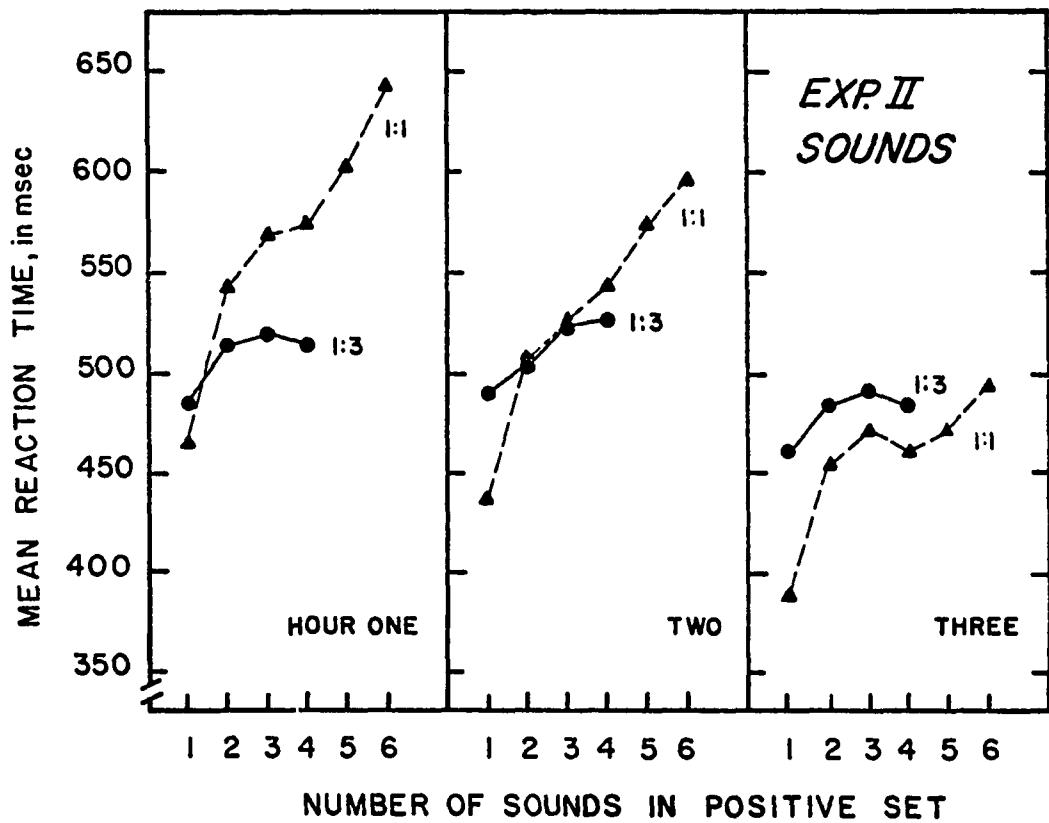
Fig. 5. Between-subject comparisons of the Sounds 1:3 and Sounds 1:1
Conditions of Experiment II, separated by hour.

structions as the subjects of Experiment II, but were encouraged to work
faster than they had previously done. [The overall error rate of the
Sounds 1:1 Condition was 5.2%, considerably higher than that of the exper-
ienced subjects working on the Sounds 1:3 Condition in Experiment I. The
error rate of the Extension 1:1 subjects, responding faster than before,
was also 5.2%.]

The data obtained from these subjects were extremely stable, show-
ing no effects of practice on either slopes or intercepts. The mean
reaction times across all subjects for all four sessions provide the
clearest estimate of performance. The mean latency, irrespective of set
size, was 77 msec. faster than that of the Sounds 1:1 subjects of Experi-
ment II (439 vs. 516 msec.). However, the slope for set sizes 2-6 was
identical to the Sounds 1:1 results -- 19 msec. per sound, accounting for
99.0% of the total variance (slope for sets 1-6 was 25 msec./sound, pre-
dicting 93.3% of the variance). The obtained latency for the one-item
Positive Set was 39 msec. faster than the time predicted from the linear
regression through sets 2-6. The comparable discrepancy for one-item
trials in the Sounds 1:1 Condition of Experiment II was 50 msec.

The data from the Sounds 1:1 and Extension 1:1 Conditions relates
to questions 2, 3, and 4, as posed at the beginning of this chapter.
Question 2 asks why the response times to four sounds were faster than
the times for three sounds. The absence of this drop for the Sounds 1:1
Conditions, in conjunction with the chunking effects observed in the Type
1:3 Condition, verifies the conclusion that the four-sound drop was an
artifact of the stimulus materials, resulting from the two-set discrimin-
ation between stimuli and Null Set items.

Question 3 asks whether the slope of the Sounds Condition would be
the same for stimulus materials lacking homonyms. The answer is simply
yes, the slope was unchanged. We are not able to conclude whether or not
the intercept is affected by homonyms, due to the equivocal findings in
Figure 5.

Question 4 asks, "Why is the average reaction time to one item in
the Looks Condition faster than the time for one sound in the Sounds Condi-
tion?" Although we are still in doubt as to the cause for one-item trials
producing latencies which are discontinuous (faster) with the other latencies,

we recognize that the fast reaction time is not restricted to the Looks
Condition. Subjects may not need to use a memory comparison stage of pro-
cessing to make a judgment involving a single target item.

Having dealt with the first four questions, we are now prepared to
consider question 5 -- was the slope of the Looks Condition affected by the
presence of homonyms? Experiment III was designed to answer this question.

EXPERIMENT III

Experiment III was designed to provide a between-subjects compari-
son of Looks and Sounds instructions for materials without homonyms.
Therefore it was run concurrently with Experiment II, using five new sub-
jects who were selected in the same manner as those in the second experiment.
The trial tape was identical to the Sounds 1:1 Condition of Experiment II,
except the instruction was changed to read, "LOOKS THE SAME." Subjects
were given the further instructions to try not to pronounce the to-be-
memorized items as they appeared, but to concentrate on their form and
letter configuration. Each subject served for only one hour, after which
he was questioned concerning rehearsal. All five subjects reported
being unable to prevent silent naming of the items, and most subjects .
felt they were rehearsing at least part of the time, having some uncer-
tainty about what they had been doing.

The results of Experiment III are displayed on the left side of
Figure 6. Plotted with these results are the data from Hour One of the
Sounds 1:1 Condition of Experiment II and Series One of the Looks Condition
of Experiment I. The two Looks Conditions are nearly parallel except for
the dip in Experiment III with set size four. The slope for Experiment III
was 42 msec./item, only slightly below the 45-msec. value which best repre-
sented the data of Experiment I. The minor difference in intercept can be
attributed to the fact that subjects in Experiment III were serving their
first hour, while the data from Experiment I has collapsed latencies from
subjects in their second, third, or fourth hour of responding. The results
of the Sounds 1:1 Condition of Experiment II were quite different from the
Looks Condition of Experiment I, the best-fitting slope for the Sounds Condi-
tion being less than 24 msec./sound, calculated across the sets of 2-6 items
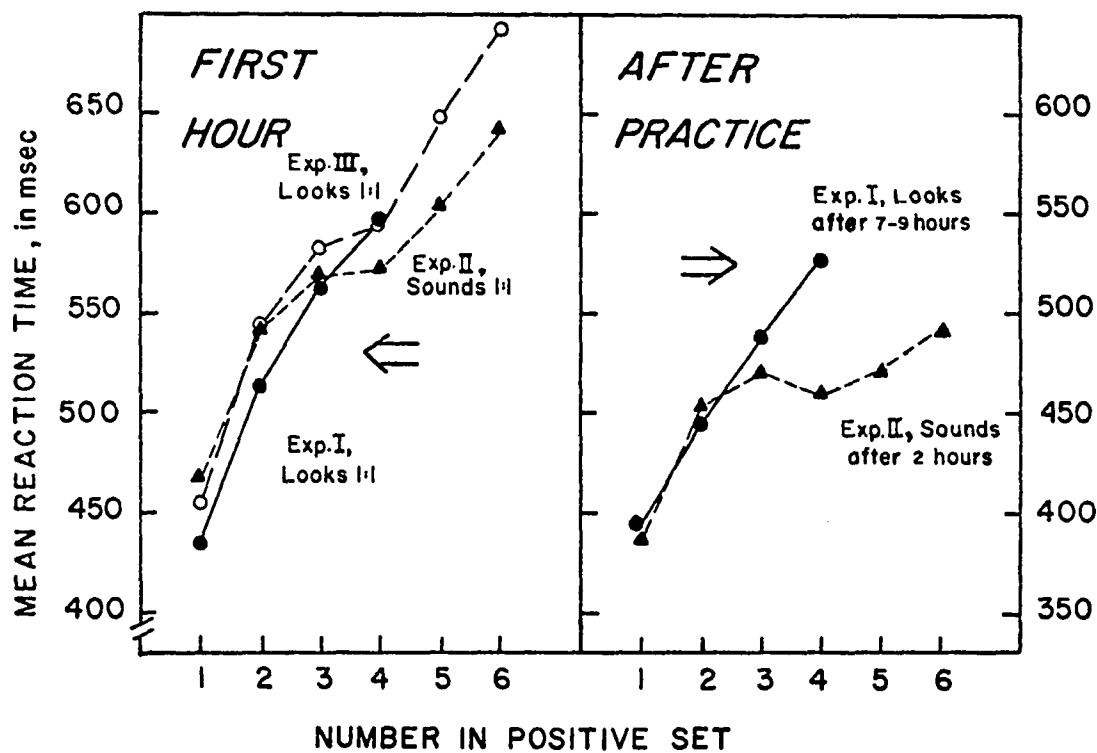
Fig. 6. A comparison of reaction times for the Sounds 1:1 Condition
with the Looks Conditions of Experiments I and III, with and
without practice.

(accounting for 94.1% of the variance). A further demonstration of the
departure of the Sounds vs. Looks data is apparent from the right side of
the figure, which illustrates the effects of practice. The steeper slope
represents the data from Series Three of Experiment I, incorporating perform-
ance after seven, eight, or nine sessions (the variation being due to the
different sequences for serving in the three conditions of Experiment I).
The shallower slope displays data from the Sounds 1:1 Condition for those
subjects working their first hour at this condition, but after two sessions
at other conditions. The slope for points 2-6 was only 8 msec./sound.
Once again we note the one-item reaction time for the Sounds 1:1 Condition
appears to be discontinuous with the rest of the Sounds function, but the
same as the one-item response time of the Looks Condition. Despite this
minor similarity between the Sounds 1:1 and Looks Conditions, Figure 6
establishes that the Sounds 1:1 Condition differs from the Looks Conditions
of both experiments.

In answer to question 5, the small differences between the Looks
Conditions of Experiments I and III suggests that the former was relatively
uninfluenced by either the presence of homonyms or the other conditions of
that experiment.

While several puzzles have been resolved, two remain. (1) Why were
the slopes of the Sounds Conditions of Experiment II flatter than those
of Experiment I? (2) If memory is solely auditory, why did the Looks
instruction of Experiment III produce a steeper slope for the same mater-
ials as used in the Sounds 1:1 Condition? This last question opens again
the possibility of two forms of memory. In summary, Experiments II and III
have resolved several difficulties in the interpretation of Experiment I.
In the process, the above two questions arose. Chapter V will resolve
these difficulties so that we may return to our original question about
the form of memory codes.

CHAPTER V

RATIONALE FOR SEQUENTIAL ANALYSIS

Our primary interest to this point has been to apply the Sternberg
paradigm to the problem of identifying the nature of those memory traces
operating in short-term recognition memory. Now we direct our attention
to the method which has served as a tool for discovery. We ask, "Are the
assumptions of a serial exhaustive search justified?" Specifically, we
shall test one implication of the serial exhaustive search model -- that
the slope should be unaffected by sequential patterns of responding.

The rationale for this prediction was set forth in Chapter I.
Three categories of sequential tests were identified: (1) Non-Repetitions
(NR), referring to a change from either a Yes to a No response or vice
versa, (2) Equivalence repetitions (EQ), where a response is the same as
on the previous trial, but the stimulus has changed, and (3) Identity
repetitions (ID), where both the stimulus and the response are the same.
Differences in intercepts for the three types of sequential responses
will not be damaging to the search model, since the Additive Factors
Method suggests these effects are produced by other stages of processing.
However, differences in slopes among these three response categories would
imply the sequential properties of responding affect the memory comparison
stage. Consequently, we will concentrate on differences in the slope
parameter.

Since the probability of a Yes response was constant ($p = .50$), the
relative distribution of IDs and EQs must necessarily vary with the size
of the Empirical Positive Set. With one item, all Yes repetitions are
IDs. With two items half will be IDs, half EQs. The extreme case of 12
distinct items in the Empirical Positive Set (or four sounds) leads to an
overall probability for an ID repetition of $(1/2)(1/12) = .042$ (therefore
the probability of an EQ would be the complement, or .458). Increasing the
size of the Positive Set will increase the likelihood of an EQ Yes repeti-
tion, but decrease the likelihood of an EQ No repetition, due to fewer Ex-
clusive items in the No response pool. The opposite relationship holds
for ID repetitions.

69

Table 8 lists the slopes for the Looks and Sounds Conditions of all experiments, divided into Yes vs. No responses for all three sequential patterns. Only the best-fitting straight lines are recorded. The range for this calculation and the regression correlation of each slope are also noted. The criterion of selecting lines having the highest correlation coefficient serves as a preliminary organizing principle. All cited values will employ this criterion unless noted otherwise. However, we will amend this criterion later in an attempt to establish the most consistent time estimation across experiments for similar instructions. [Appendix G lists the mean reaction times as a function of set size and sequential categories for those conditions listed in Table 8.]

Typical patterns of results can be seen in Figures 7 and 8, both displaying Yes responses, by sessions. Figure 7 shows the Looks Condition of Experiment I while Figure 8 displays the Sounds 1:1 Condition of Experiment II. First we should note that IDs usually produce the flattest slopes, EQs the steepest slopes. The NRs, which constitute half the total responses, provide the most stable slopes. For example, in Experiment I the Looks Yes NRs had slopes of 45, 45, and 49 msec. per item, for Series One, Two, and Three, respectively. The No NRs for these same series resulted in slopes of 41, 45, and 47 msec. per item. In contrast, Yes ID slopes ranged from 18 to 38 msec. for these series and No IDs ranged from 47 to 105 msec. per item -- the latter values being the only exceptions to our generalization that ID responses produce the flattest slopes.

The importance of the stability of the NR slopes is most apparent in Figure 8. The steep slope for Hour One subjects with all responses combined was previously perplexing (see Figure 5). Now it is apparent that much of that effect was contributed by EQ responses. The Yes NR slopes vary within 10 msec. The No response NRs assume best-fit values of 21, 27, and 3 msec. per sound for the three hours -- only the last value being difficult to explain. Therefore two sequential components contributed to the previously noted shallow slope of Hour Three subjects -- a shallow NR slope and a curvilinear ID function which involved shorter latencies with larger Positive Sets.

TABLE 8

REACTION TIME SLOPES AND REGRESSIONS BY SEQUENTIAL

CATEGORIES FOR ALL LOOKS AND SOUNDS CONDITIONS

Response			LOOKS		SOUNDS 1:3		SOUNDS 1:1	
			I	III	I	II	II	Ext.
Yes								
	ID	b	32.7	35.8	40.0	12.7	13.5	17.0
		s	1-4	1-6	1-3	1-4	1-6	1-6
		r	.971	.979	.997	.913	.796	.948
	EQ	b	53.5	47.2	.40.0	24.0	32.9	20.6
		s	2-4	2-6	1-3	1-3	2-6	2-6
		r	.976	.956	.996	.990	.988	.987
	NR	b	51.5	31.9	21.7	20.0	15.8	24.7
		s	1-4	2-6	1-4	1-3	2-6	2-6
		r	.976	.983	.970	.918	.998	.994
No								
	ID	b	77.0	-11.5	33.9	-4.4	19.9	12.4
		s	1-3	2-6	1-4	1-4	1-6	1-6
		r	.995	-.658	.894	-.516	.806	.561
	EQ	b	38.1	35.9	25.5	14.5	21.1	16.1
		s	1-4	1-6	1-3	1-3	1-6	2-6
		r	.995	.918	.999	.947	.849	.952
	NR	b	44.4	42.5	22.5	6.5	17.2	17.8
		s	1-4	1-6	1-3	1-3	2-5	2-6
		r	.981	.943	.990	.866	.999	.978

b = slope; s = Positive Sets used in calculation; r = correlation about slope
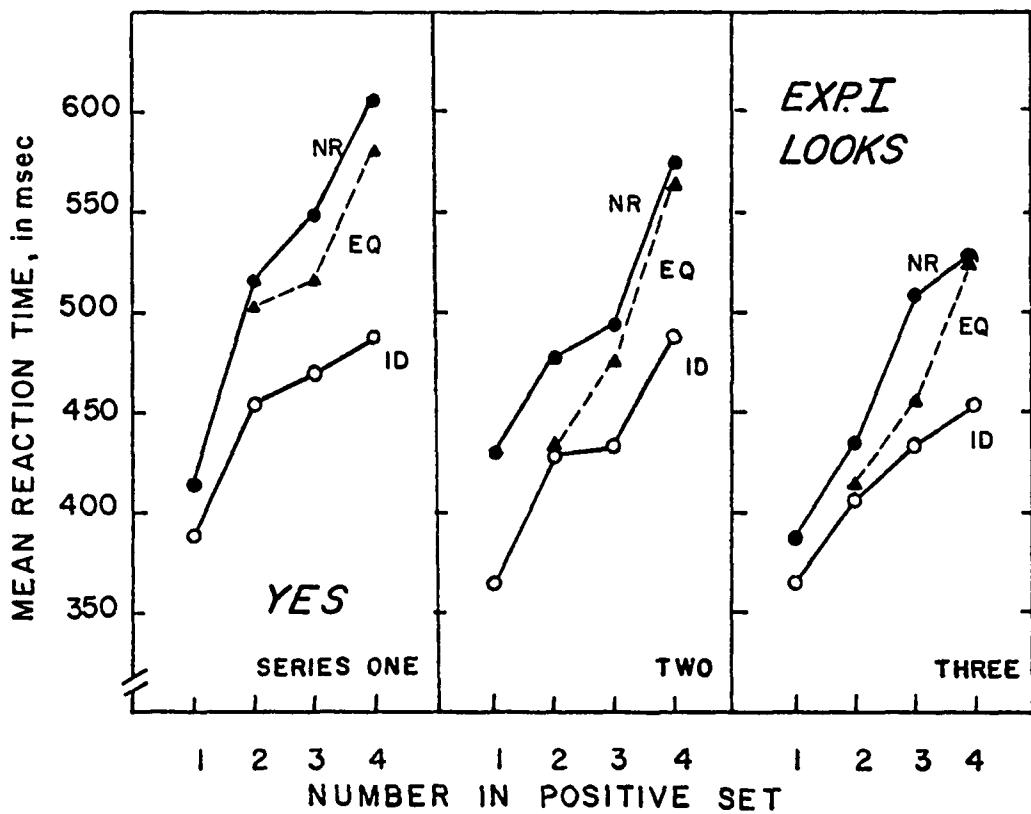
Fig. 7. Sequential effects for Yes responses in the Looks Condition of
Experiment I, displayed by series. NR represents Non-repeti-
tions, EQ refers to Equivalence repetitions, and ID refers to
Identity repetitions.
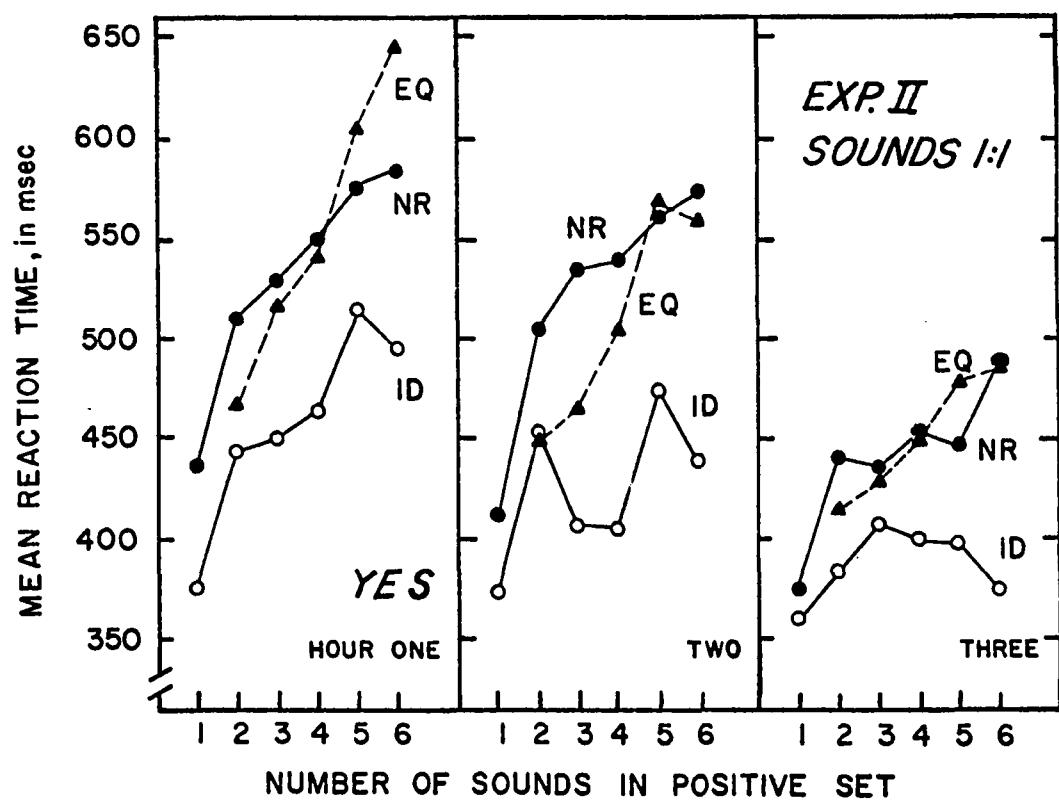
Fig. 8. Sequential effects for Yes responses in the Sounds 1:1
Condition of Experiment II, displayed by hour.

The lack of agreement among the three sequential response categories using the slope parameter to characterize the results of any condition can be neither overlooked nor simply explained from the present findings. Such sequential effects are clearly contrary to the predictions from the serial exhaustive search model. The variables of the present study do not appear to provide a clue for understanding these effects, since different patterns emerged across those conditions which have otherwise seemed comparable. At best, the present study has merely called attention to a phenomenon which requires explanation.

IN SEARCH OF A SINGLE COMPARISON TIME

The stability of NR slopes suggests that this component of the reaction time data be used to estimate the average comparison time. The comparison time is thought to represent the time to match the encoded version of the test item with a single memory trace. Without partitioning of sequential effects, the overall slopes in the Sounds Conditions ranged from a high of 30 msec. per sound in Experiment I to a low of 17 msec. per sound in Experiment II (Sounds 1:3). Do the NR slopes more nearly converge on a single comparison time across experiments?

Examining the NR slopes for both Sounds 1:3 and Sounds 1:1 Conditions in Table 8, the time of about 20 msec. emerges as a good estimate for Yes responses, being slightly underestimated for the Sounds 1:1 Condition, of Experiment II, but compensatingly overestimated by the Extension 1:1 data. This time estimate is also a good fit for No responses in Experiment I and the two Sounds 1:1 Conditions. Can these estimates be improved, reducing their range? First we should recall that the single criterion for listing those slopes which appear in Table 8 was the magnitude of correlation. Let us now consider other slope calculations computed over alternative ranges.

The Sounds 1:1 results are most susceptible to modification. Calculating slopes across the full range of memory sets, the Experiment II slopes become 24.5 and 23.2 msec. for Yes and No responses, respectively. The predictability has not fallen drastically by this alteration ($r = .925$ and .944, respectively). The Yes slope for Extension 1:1 is an acceptable

24.7 msec. per sound. Once again resorting to a computation across all
six sets, the No slope becomes 23.3 msec. (r = .958). The agreement
among slopes in these two Sounds 1:1 Conditions continues to be remarkable.

Disregarding the 7 msec. No response slope in the Sounds 1:3 Condi-
tion of Experiment II (which had given us trouble in Chapter IV), the re-
maining seven estimates of the auditory comparison time have a mean of
22.8 msec. and a high-low range of less than five msec. The aberrant
7-msec. slope remains a mystery, perhaps related to the extremely slow
reaction times for No responses during that condition.

A detailed analysis of sequential effects has verified that the
Sounds Conditions, which we expected to be alike, but which differed
between Experiments I and II prior to a partitioning of sequential cate-
gories, were in fact the same for NR latencies. Thus we have resolved
the first puzzle resulting from Experiment II (as stated at the close
of Chapter IV). The second difficulty arising from the additional
experimentation involved the discrepancy in slopes comparing the Sounds
1:1 Condition of Experiment II with the Looks Condition of Experiment III.
Both conditions used the same materials, but the slope of the Looks Condi-
tion was greater. Again, looking only at the NR slopes, this discrepancy
evaporates. While the slope of the Looks Condition of Experiment I re-
mains unchanged by the sequential partitioning, the slope of the Looks
Condition of Experiment III was reduced to a value consistent with those
obtained in the Sounds Conditions (i.e., about 23 msec.). Specifically,
the slope for combined Yes and No responses was 23.9 msec. per item
(Yes slope = 27.3 msec., r = .982; No slope = 20.4 msec., r = .907).
These calculations exclude those trials displaying either a single item
or six items. [The exclusion of set size six is reasonable for this con-
dition since these trials produced twice as many errors as those with five
items -- 4.5% vs. 10.5%.] By contrast, the flattest NR slopes for the
Looks Condition of Experiment I were 45 (Series Two) and 41 msec. (Series
One) for Yes and No responses, respectively. Therefore, we conclude the
results of the Looks Condition of Experiment III reflect the same type of
processing used in the Sounds Conditions. By restricting our attention
to NR slopes, the direct comparison of different conditions and of separate
experiments suggests we are measuring the same process -- a serial, exhaus-
tive search through auditory memory codes.

CHAPTER VI

CRITERION III FOR THE AUDITORY HYPOTHESIS

We return to our initial strategy of determining which hypothesis
best accounts for the results of Experiment I. The analysis of errors
and the supplementary evidence discussed in Chapters IV and V have served
to bolster our confidence in the Auditory Hypothesis. We have already
noted the fast absolute reaction times, the shallow slope, and the low
intra-set variability of the Sounds Condition, providing evidence for
the Auditory Hypothesis (consistent with Criteria I and II). Criterion III
requires that the Auditory Hypothesis reorder the trials of the Looks and
Means Conditions into Auditory Positive Sets such that the newly calculated
reaction time slope for each condition be the same as that obtained for
the Sounds Condition. Additionally, the resulting intra-set variability
should be less than those values reported in Table 2, where these conditions
were ordered according to alternative hypotheses. Encouragement to apply
Criterion III to the Auditory Hypothesis derives from the quantitative
similarity of auditory confusions comparing the Looks and Means Conditions
(Table 5). Since we observed nearly identical numbers of false positive
homonym errors for these two conditions, we have reason to believe that
the memory code functioned the same way in each and every condition of
Experiment I.

The strategy for reanalysis of the Looks and Means Conditions is
similar to the initial proposal for ordering trials, as outlined briefly
in Chapter III. Once again, the weighted response times for all six
subjects were combined for each trial. All trials which have the same
number of sound units are grouped together. Since these groupings are
based on a theoretical interpretation of equivalence (as opposed to using
the objective stimulus variables of C or D), it is imperative that these
classification rules be clearly stated.

The classification was based on a modified form of the Auditory
Hypothesis. It may be termed the Sound-Tag Model. The model proposes
that memory traces are isomorphic with the sound of a spoken word. When
the stimuli have unique sounds (no homonyms), we conceive that subjects
make a serial exhaustive search through a memorized list of sounds.
When homonyms are introduced and the different versions of a sound belong
to separate categories, it is not sufficient to store the sound alone.
We assume that the subject stores the sound, but in addition, he also
stores a "tag." A tag can be conceived as a private code which tells
the subject which version of the homonym belongs in the Empirical Positive
Set. Just as it takes time to compare a test to a memorized sound, we
assume it takes time to read the tag. The comparison time for a tagged
sound will be longer than the time to check a sound without a tag. The
comparison time for the tag can be calculated by simply subtracting the
slope across trials requiring tags from the slope across those trials
where the sound alone is sufficient for correct classification. If
there is no slope difference between these sets of trials, we have no
evidence that a tag was used.

We further need to consider two possibilities. Suppose that one
trial contains the set "1, 2, 4" whereas a second trial contains the
set "ONE, 2, FOR." The instruction is "looks the same." Both sets con-
tain three sounds. But the first set involves only one Type -- Arabic
numbers -- whereas the second set involves a mixture of Types. That a
subject's performance reflects a recognition of the difference between
these sets was suggested by the error analysis. False positives to incorrect
Arabics occurred more often on those trials which displayed all Arabics.
The Sound-Tag Model separates the two sets by proposing that trials involv-
ing homonyms which have a common feature (as when all items are Arabics)
may require a single tag for the entire set, rather than a separate tag
for each sound. This type of trial will be termed a "one-tag" trial,
since we will assume the subject needs to store only a single tag regard-
less of the number of memorized sounds. When there is a mixture of Types,
we assume that the subject must attach one tag to each and every sound.
These trials will be termed "one tag/sound" trials.

Ideally, the one tag/sound category would include only those trials where there are an equal number of Types and sounds. This would insure that each sound had a unique tag. Otherwise, subjects might cluster sounds and tags. For example, the Positive Set of 1, 2, FOR, ATE might be organized hierarcnically with two binary classifications: first a search through the two tags (Arabic vs. non-numbers), then a search through the two sounds using the selected tag. Williams (1971) has demonstrated that subjects can use this hierarchical strategy in the memory search task. Since the present study used up to four sounds, but only three different versions of a sound, the ideal exemplars for the one tag/sound category were not available. In general, the one tag/sound category will include fewer Types than sounds for the larger Positive Sets.

The definition of a tag differs for the two instructions in Experiment I. For the Looks Condition a tag is needed for each of the three Types, while in the Means Condition a tag denotes one of the two semantic forms of a sound. Trial 25, displaying the items ONE, 2, illustrates this point. For the Looks Condition, this trial would be categorized as a one tag/sound trial. However, in the Means Condition, it would be categorized as a one-tag trial since both sounds use the same code (i.e., number versions). Table 9 lists those trials assigned to each category for the present analysis.

The predicted slope for the one-tag trials differs from that of the one tag/sound trials. If a single tag is sufficient, a single time increment is expected, since the search through additional sounds would take the same comparison time as in the non-homonym (therefore no tag) situation. The predicted slope is parallel to the non-homonym slope, but has a constant increment associated with the tag. We expect the slope of the one-tag trials to be about 23 msec. per sound -- the average value for the various Sounds Conditions as calculated in Chapter V. However, the intercept for the one-tag trials should be higher, reflecting the time to examine one tag.

The one tag/sound trials should have a steeper slope than the one-tag trials, since each comparison time now includes the tag-check time added to the sound-match time. We assume the tag would be checked regardless of the outcome of the sound match, consistent with the concept of an

TABLE 9

LIST OF TRIALS FROM THE LOOKS AND MEANS CONDITIONS

OF EXPERIMENT I WITHIN EACH SOUND-TAG CATEGORY

Theoretical Positive Set	Looks Condition	Means Condition
One Item	1,2,3,4,5,6,7, 8,9,10,11,12	3,6,9,12
	Complete Triplet	Complete Triplet
One Sound	31,32,33,34	31,32,33,34 18,21,24*
Two Sounds	--	50,54,55*

	One-Tag	Tag/Sound	One-Tag	Tag/Sound
Two Sounds	26,29,30	25,27,28	25,26,28 29,30	27,40
Three Sounds	36,38	42,43	36,38,42, 49	43,48
Four Sounds	44,45	46,47	44,45,46	47

* All three versions of the triplet were inferred to be correct, although not all versions were displayed. Trials 31-34 displayed all three forms of the homonym triplet.

exhaustive search. If we assume the time to check a tag is equal to the time to match a sound, the slope of the one tag/sound trials would be twice that of the one-tag trials. Wattenbarger's finding of nearly a 2:1 slope differential for Physical Identity trials (presumably requiring a tag for upper vs. lower case) relative to the Name Identity trials supports this "ball-park" estimate.

There are two ways we can estimate the time required for a tag check. The first is to subtract the intercept of the sounds-alone (no tags) trials from the intercept of the one-tag trials. The other method subtracts the slope of the one-tag trials (or sounds-alone trials, since the slopes should be identical) from the one tag/sound trials, since the slope measures the time for a one-item comparison (sound comparison vs. sound + tag comparison).

In partial summary, the Sound-Tag Model predicts: (1) The slopes for the one-tag and one tag/sound trials will be linear. (2) One-tag trials should have the same slope as the Sounds Conditions, namely about 23 msec. per sound. (3) The one tag/sound trials should produce a steeper slope than the one-tag trials. (4) The lines defining each function should converge at a Theoretical Positive Set size of one sound unit, since at this point each function represents the time to search one sound plus one tag.

The overall mean trial times were first partitioned among the two sound-tag categories, and then grouped according to the size of the Theoretical Positive Set. Figure 9 displays the results for both Looks and Means Conditions. Let us first consider the data from the Looks Condition (left side of figure). With one sound in the Theoretical Positive Set, trials were divided into those displaying a single item (open square) vs. those displaying all versions of a homonym triplet (open circle). They differ by approximately 50 msec., indicating once again that the one-item trials may involve different processing. Part of this difference may be due to marked differences in the probability of a particular Positive Set stimulus (p = .50 for one item vs. .17 for homonym triplets). Theios, Smith, Haviland, and Traupmann (1971) have shown that stimulus probability plays a major role in determining the reaction time during a memory search task. In addition to the contribution of stimulus probability, the distribution of sequential response categories differed (all Yes response repetitions were IDs for one-item trials vs. 1/3 IDs, 2/3 EQs for one-sound trials).
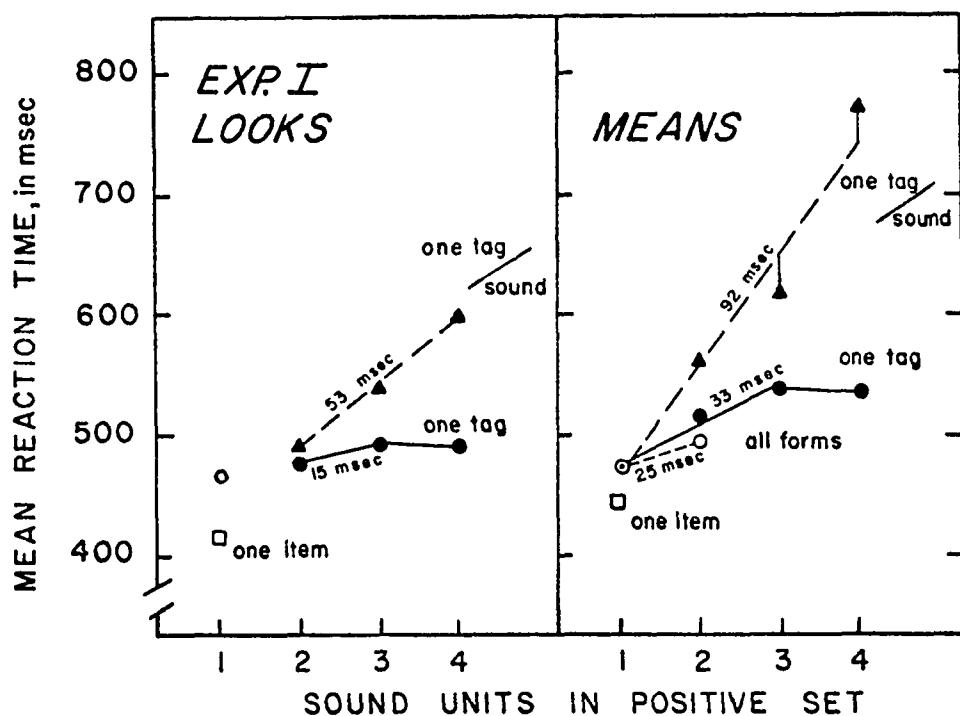
Fig. 9. Reorganization of trials from the Looks and Means Conditions of Experiment I according to the Sound-Tag Model, displayed in terms of the size of the Theoretical Positive Set and separated according to the presumed number of tags needed to specify the correct stimuli.

The one-tag trials result in a function similar to that of the Sounds Condition, including the drop with all four sounds. The increment from set size two to three was 15 msec., somewhat below the slope from the Sounds Condition. [This value was derived from weighted means, rather than the unweighted values in Apprendix B.] This curve is markedly different from the one tag/sound function which has a slope of 53 msec. per sound. The two functions converge near the one-sound location, as predicted.

Data from the Means Condition were similar. The one-item trials were faster than the one-sound trials. However, the comparison comparable to that with the Looks Condition produced only a 16-msec. difference between these two types of trials, whereas this difference had been 50 msec. in the Looks Condition.

With the Means instruction, there were a set of trials where all three versions of a homonym triplet could be correct for both one and two sound units in the Positive Set (labelled "all forms" on Figure 9). The difference between the one- and two-sound trials was 25 msec., very close to the expected comparison time of 23 msec./sound. This calculation compares those trials where all forms were inferred to be correct (e.g., displaying 1, WON implies 1, ONE, WON), although some were not displayed. Displaying all members of the triplet (e.g., 1, ONE, WON) in the one-sound case resulted in a 9-msec. faster response time relative to the inferred one-sound trials.

The one-tag curve also approximates the slope from the Sounds Condition. The slope of 33 msec. per sound is somewhat steep, but the presence of greater numbers of faster ID responses for the one-unit trials may have biased this estimate. The difference between set size two vs. three (less susceptible to repetition imbalance) was, encouragingly, 24 msec.

The one tag/sound trials produced the very steep slope (92 msec. per sound) on the right side of Figure 9. The steepness was contributed largely by the point representing four sound units in the Theoretical Positive Set, which, unfortunately, was determined by a single trial (Trial 47). A more reasonable estimate of the one sound/tag function can be obtained using sets two and three, which produce a slope of 57 msec. per sound, similar to the

one tag/sound slope for the Looks Condition (54 msec.). As predicted by
the Sound-Tag Model, the line through these two points intersects the one-
tag function, when determined by sets of two and three sounds only, at the
location of set size one. This intersection point, which is higher than
the plotted one-sound point, is 45 msec. slower than the one-item trials.
The 45-msec. value more nearly agrees with both the one-sound vs. one-item
difference of the Looks Condition (50 msec.), and the obtained vs.
predicted discrepancies for the one-item trials in the Sounds 1:1 (50 msec.)
and Extension 1:1 Conditions (39 msec.), as reported on page 65.

As noted earlier in this chapter, the tag-check time can be esti-
mated in two ways: by comparing intercepts for the one-tag vs. no-tag
trials, and by comparing slopes for the one-tag vs. one tag/sound trials.
Since the latter comparison can be made for both the Looks and Means
Conditions, let us complete these calculations first. In the Looks Condi-
tion, the slope differential was 53 - 15 = 38 msec. This value may over-
estimate the tag-check time due to the underestimation of the one-tag
slope (expected to be 23 msec., which would yield a value of 30 msec.
for the tag-check time). For the Means Condition the most reasonable
estimate compares the slopes across only sets of two and three sounds.
This difference is 57 - 24 = 33 msec. per tag.

The comparison of intercepts can be completed only within the Means
Condition. The difficulty with the Looks Condition is that we can not
calculate a slope (and hence derive an intercept) for the no-tag trials,
since there is only one set size where all three versions of a homonym
triplet are correct. Even in the Means Condition the no-tag slope can
only be estimated by piecing together two components of the difference
of interest. We wish to determine the difference in reaction time between
trials requiring one tag and trials where all members of the homonym triplet
are displayed. We derive this value by getting the difference between
one-tag and no-tag slopes where the latter trials have all homonym versions
inferred. To this we add the difference in latency between trials where
all versions are inferred vs. all versions displayed. Since the slopes of
the one-tag (through sets 2 and 3) and of the no-tag or all-forms trials
(through sets 1 and 2) are virtually identical, we can measure the differ-
ence between the two functions at set size 2. This difference is 19 msec.

The difference between the function where all forms of a sound are inferred vs. where all forms are displayed is 9 msec. Therefore, the difference between the one-tag and no-tag trials is estimated to be 19 + 9 = 28 msec. We assume this estimate reflects the time to search one tag. Considering this 28-msec. estimate along with the above estimates, we arrive at a tag-check time of about 30 msec., compared to about 23 msec. to search a single sound representation. These estimates do not depart sufficiently to reject the conclusion that the tag-check time is equal to the sound-match time.

In conclusion, the Sound-Tag Model has effectively partitioned the data into two functions: one conforming to the function obtained in the Sounds Condition -- thereby satisfying Criterion III -- and the second derived from the former function by adding the time for tag checks. It was this one tag/sound component which contributed heavily to the steep slopes of the Looks and Means data in the results of Experiment I. We now have good reason to believe that subjects use an auditory code regardless of the instructions. Thus our Sound-Tag Model has synthesized the apparent discrepancy among conditions in Experiment I.

The intra-set variability, calculated for each function separately, must necessarily be reduced in terms of the components of analysis. The actual calculation of variability is not practical due to the small number of exemplars at each set size for each category (see Table 9). Nevertheless, it is intuitively clear that the variance has been partitioned into two segments, reducing the range of reaction times within each segment. In summary, the Sound-Tag Model allows us to reasonably satisfy Criterion III, leading to the conclusion that short-term recognition memory relies on an auditory form of memory code.

CHAPTER VII

CONCLUSIONS

We have probed short-term recognition memory with a reaction time
procedure which claims to reveal some temporal aspects of memory-related
central processes. Our metaphorical excursion through three experiments
began with three conjectures about the properties of memory. We enter-
tained the possibilities that different forms of memory codes function
during a word recognition task. The three forms were assumed to preserve
the auditory, visual, or semantic features of the Positive Set items.
The evidence strongly points to the conclusion that short-term recognition
memory codes are basically auditory for verbal materials. Now the time
has come to review this evidence and to examine its soundness.

Six facts unequivocally support the conclusion that memory codes
are auditory in the Sternberg stimulus classification task:

1. The fastest reaction times occurred in the Sounds Condition
of Experiment I. This result is even more impressive given that this
condition required the memorization of the largest number of physically
distinct visual patterns. The probability that any one of the Positive
Set items could occur as a test item was lowest in this condition. This
latter finding stands in direct contradiction to the claim of Theios et
al. (1971) that the stimulus probability is the critical variable deter-
mining reaction times in the memory search paradigm. Although the Sounds
Condition was at the greatest disadvantage in terms of both the number of
to-be-memorized stimuli and the probability of a given stimulus appearing
as a test item, it produced the fastest response latencies.

2. The slope, representing the increase in reaction time with an
increased number of items in the Theoretical Positive Set, was shallowest
in the Sounds Condition of Experiment I. If we accept the proposition
that these increments reflect the operation of a memory-related stage of
processing, we must conclude that attention to the auditory features of

the to-be-memorized items facilitates such processing, relative to emphasizing the visual or semantic features of the stimuli. This conclusion is essentially independent of the assumptions of Sternberg's serial exhaustive search model. That is, the shallow slope of the Sounds Condition, relative to the other two conditions, is evidence for auditory memory codes whether the memory search is serial or parallel, exhaustive or self-terminating. The only necessary assumption for this conclusion is that the increments in reaction time with larger Positive Sets reflect the use of memory codes, rather than some other stage of processing.

3. There was both less total variability and less intra-set variability within the Sounds Condition of Experiment I than in the other two conditions. Therefore, the opportunity to organize the to-be-memorized items according to their auditory properties provides the most efficient basis for the memory-dependent response classification. The Sounds Condition was again at the greatest disadvantage since it involved the greatest discrepancy between the set of displayed stimuli and the set of potentially correct items. That is, the Sounds Condition provided the least stimulus (D variable) to target (C variable) compatibility. During the Looks instructions, the subject simply gave a Yes response to the same items as in the displayed set on all 48 trials. By comparison, only 4 of the 48 trials permitted this direct stimulus to target correspondence in the Sounds Condition. For the remaining 44 trials the subject had to determine which additional (non-displayed) items should also receive a Yes response. Despite this added complexity, the subjects gave their most consistent performance under the Sounds instruction.

4. The analysis of errors revealed an overwhelming tendency to incorrectly respond Yes to a homonym of a Positive Set item in the Looks and Means Conditions. By contrast, there was no substantial evidence of visual or semantic confusions. The evidence from false positive responses alone was so strong that it warrants the conclusion that memory codes retain auditory properties. Therefore, we need not rely solely upon latency measures to verify that memory is auditory.

5. The slope of the Sounds Condition was independent of the type of stimulus materials, within the range tested by the present research. This includes both with and without homonyms present and for numbers as well as three-letter words (Sounds 1:3 vs. Sounds 1:1). This result contradicts the claim that the stimulus probability is the critical factor. Stimulus probability varied over a 3:1 range in comparing the Sounds 1:3 with the Sounds 1:1 Conditions, since these conditions represented a 3:1 vs. a 1:1 ratio of physically distinct homonym versions per sound. This result is most reasonable if a serial exhaustive search model is assumed, but it is equally sound for other models, provided that the memory code is auditory.

6. The same slope was obtained with instructions emphasizing other than auditory features of the stimuli, according to the Sound-Tag interpretation. This finding includes both results of Experiment I (Looks and Means Conditions) and of Experiment III (Looks without homonyms present). This result adds to our confidence that the theoretical basis for this prediction is well founded (i.e., that subjects engage in a serial exhaustive search through auditory codes). Even with a task designed to favor some other encoding strategy (visual or semantic memory), subjects persist in using auditory codes -- despite the fact that the auditory codes lead to homonym errors.

A possible objection to the conclusion that short-term recognition memory involves solely auditory codes is that memory is multidimensional. This position has gained support from experiments in which it has been shown that subjects can respond to auditory, visual, or semantic features, as directed. On this basis alone, the multidimensional position is indistinguishable from the present position which holds that the primary mode of encoding is auditory with tags appropriate to other features attached to the name codes. While stimulus classifications made briefly after the onset of a to-be-memorized item may benefit from visual similarity, the present study has only found evidence of auditory properties for time intervals ranging from 5 to 23 seconds after the onset of the to-be-memorized list. Within this interval, we have discovered memory-related processes which require under 25 milliseconds per comparison. To consider that semantic features are available some 200-300 msec. later than auditory

properties (as in Shulman's results, 1970), permits a large amount of time for other cognitive operations to take place on the auditory code. We have made the strong assumption that elapsed time implies elapsed processing, an assumption which has not been disconfirmed in the present findings.

While the present findings are not inconsistent with the multi-dimensional trace models, they do not provide evidence for any other properties than auditory ones. The results are compatible with the unitary code theory, a position which makes stronger (i.e., easier to disconfirm) predictions. It is not clear what evidence could disconfirm the more general versions of a multidimensional theory.

In tempering the conclusion that memory codes are auditory we should note once more that the present research used only verbal materials, and that we are unable to state whether this acoustic-like memory code follows more closely the perception of speech, the articulation of words, or the physical properties of sound waves.

While the conclusion that memory is auditory for verbal materials does not rest on any single fact, nor on the frailty of the totality of facts, we are compelled to accept this conclusion given the convergence of evidence from diverse sources. The conclusion relies on neither a single dependent variable (absolute latency, slopes, and errors have been considered), nor on one set of theoretical assumptions (not just the serial exhaustive search model). Nevertheless, the entire set of findings form a consistent picture when interpreted according to the serial exhaustive search model proposed by Sternberg (1969). It is this model which permits the straight-forward interpretation of the magnitude of the slope (i.e., about 23 msec. per sound) as representing the time to make a single memory comparison. Furthermore, this model should be credited with both prompting these experiments and parsimoniously organizing their results.

REFERENCES

Atkinson, R. C., Holmgren, J. E., and Juola, J. F. Processing time as influenced by the number of elements in a visual display. Perception and Psychophysics, 1969, 6, 321-326.

Bertelson, P. Serial choice reaction-time as a function of response versus signal-and-response repetition. Nature, 1965, 206, 217-218.

Bower, G. A multicomponent theory of the memory trace. In K. W. Spence and J. T. Spence (Eds.), The Psychology of Learning and Motivation, Vol. I. New York: Academic Press, 1967.

Broadbent, D. Perception and Communication. New York: Pergamon Press, 1958.

Chase, W. G., & Posner, M. I. The effect of auditory and visual confusability on visual and memory search tasks. Paper presented May, 1965 at Midwestern Psychological Association, Chicago.

Cohen, G. A comparison of semantic, acoustic and visual criteria for matching of word pairs. Perception and Psychophysics, 1968, 4, 2-3-204.

Conrad, R. Acoustic confusions in immediate memory. British Journal of Psychology, 1964, 55, 75-84.

Conrad, R. Interference or decay over short retention intervals. Journal of Verbal Learning and Verbal Behavior, 1967, 6, 49-54.

Dale, H. C. A., & Gregory, M. Evidence of semantic coding in short-term memory. Psychonomic Science, 1966, 5, 75-76.

Donders, F. C. Over de snelheid van psychische processen. Onderzoekingen gedaan in het Physiologisch Laboratorium der Utrechtsche Hoogeschool, 1868-1869, Tweede reeks, II, 92-120. Translated by W. G. Koster in W. G. Koster (Ed.), Attention and Performance II. Acta Psychologica, 1969, 30, 412-431.

Keele, S. W., & Chase, W. G. Short-term visual storage. Perception and Psychophysics, 1967, 2, 383-386.

Kintsch, W., and Buschke, H. Homophones and synonyms in short-term memory. Journal of Experimental Psychology, 1969, 80, 403-407.

Kornblum, S. Sequential determinants of information processing in serial and discrete choice reaction time. Psychological Review, 1969, 76, 113-131.

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

Morin, R. E., DeRossa, D. V., & Stultz, V. Recognition memory and reaction time. In A. F. Sanders (Ed.), Attention and Performance I. Acta Psychologica, 1967, 27, 298-305.

Murdock, B. B., Jr. Short-term memory for associations. In D. A. Norman (Ed.), Models of Human Memory. New York: Academic Press, 1970.

Norman, D. A. Toward a theory of memory and attention. Psychological Review, 1968, 75, 522-536.

Norman, D. A., & Rumelhart, D. E. A system for perception and memory. In D. A. Norman (Ed.), Models of Human Memory. New York: Academic Press, 1970.

Posner, M. I. Short term memory systems in human information processing. Acta Psychologica, 1967, 27, 267-284.

Posner, M. I., Boies, S. J., Eichelman, W. H., & Taylor, R. L. Retention of visual and name codes of single letters. Journal of Experimental Psychology Monograph, 1969, 79, No. 1, Part 2.

Posner, M. I., & Keele, S. W. Decay of visual information from a single letter. Science, 1967, 158, 137-139.

Posner, M. I., & Konick, A. F. Short-term retention of visual and kinesthetic information. Organizational Behavior and Human Performance, 1966, 1, 71-88.

Posner, M. I., & Mitchell, R. F. Chronometric analysis of classification. Psychological Review, 1967, 74, 392-409.

Shulman, H. G. Encoding and retention of semantic and phonemic information in short-term memory. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 499-508.

Sperling, G. The information available in brief visual presentations. Psychological Monographs, 1960, 74, (11, Whole No. 498).

Sperling, G. A model of visual memory. Human Factors, 1963, 5, 19-31.

Sperling, G. Successive approximations to a model for short-term memory. Acta Psychologica, 1967, 27, 285-292.

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

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 May 7, 1971 at Midwestern Psychological Association, Detroit.

Swanson, J. M., & Briggs, G. E. Information processing as a function of speed versus accuracy. *Journal of Experimental Psychology*, 1969, 81, 223-229.

Swanson, J. M., & Briggs, G. E. The neglected negative set. Paper presented May 7, 1971 at Midwestern Psychological Association, Detroit.

Theios, J., Smith, P. G., Haviland, S. E., & Traupmann, J. Memory search is a serial, self-terminating process. Wisconsin Mathematical Psychology Program Technical Report No. 71-1, University of Wisconsin, May, 1971.

Wattenbarger, B. L. The representation of the stimulus in character classification. Human Performance Center Technical Report No. 22, University of Michigan, 1970.

Wickelgren, W. A. Acoustic similarity and intrusion errors in short-term memory. *Journal of Experimental Psychology*, 1965, 70, 102-108.

Wickelgren, W. A. Auditory or articulatory coding in verbal short-term memory. *Psychological Review*, 1969, 76, 232-235.

Wickelgren, W. A. Multitrace strength theory. In D. A. Norman (Ed.), *Models of Human Memory*. New York: Academic Press, 1970.

Wickens, D. D. Encoding categories of words: An empirical approach to meaning. *Psychological Review*, 1970, 77, 1-15.

Williams, J. D. Information processing in a Sternberg task using varying numbers of memory ensembles. Paper presented May 7, 1971 at Midwestern Psychological Association, Detroit.

APPENDICES

APPENDIX A

TRIAL CONTENTS FOR EXPERIMENT I

Trial Number	Contents	V	S	A	Trial Number	Contents	V	S	A
1	1	1	1	1	31	1 ONE WON	3	3	1
2	ONE	1	1	1	32	2 TWO TOO	3	3	1
3	WON	1	1	1	33	4 FOUR FOR	3	3	1
4	2	1	1	1	34	8 EIGHT ATE	3	3	1
5	TWO	1	1	1	35	ONE WON FOUR	3	3	2
6	TOO	1	1	1	36	1 2 8	3	3	3
7	4	1	1	1	38	WON TOO ATE	3	3	3
8	FOUR	1	1	1	39	1 8 ATE	3	3	2
9	FOR	1	1	1	40	WON 8 EIGHT	3	2	2
10	8	1	1	1	41	TWO TOO FOR	3	3	2
11	EIGHT	1	1	1	42	2 4 EIGHT	3	3	3
12	ATE	1	1	1	43	TOO FOR 8	3	3	3
13	1 ONE	2	1	1	44	1 2 4 8	4	4	4
16	2 TWO	2	1	1	45	WON TOO FOR ATE	4	4	4
18	TWO TOO	2	2	1	46	ONE TWO 4 EIGHT	4	4	4
21	FOUR FOR	2	2	1	47	1 TOO 4 ATE	4	4	4
22	8 EIGHT	2	1	1	48	1 ONE TOO FOR	4	3	3
24	EIGHT ATE	2	2	1	49	1 ONE 2 8	4	3	3
25	ONE 2	2	2	2	50	ONE WON EIGHT ATE	4	4	2
26	WON TOO	2	2	2	51	1 TWO TOO ATE	4	4	3
27	WON FOUR	2	2	2	52	1 4 FOR EIGHT	4	4	3
28	2 EIGHT	2	2	2	53	2 TOO 4 FOUR	4	3	2
29	4 8	2	2	2	54	2 TOO FOUR FOR	4	4	2
30	FOR ATE	2	2	2	55	FOR FOUR EIGHT ATE	4	4	2

Practice Trials

10	8	1	1	1
29'	8 4	2	2	2
37	ONE TWO EIGHT	3	3	3
51'	TWO ATE 1 TOO	4	4	3
53'	2 4 TOO FOUR	4	3	2
56	TOO 4 EIGHT WON	4	4	4
57	1 2 4 8 ONE TWO	6	4	4
58	ATE FOUR 8 ONE 2 FOR	6	6	4

* Number of visual (V), semantic (S), and auditory (A) units in Positive Set

APPENDIX B

TRIAL BY TRIAL REACTION TIME MEANS FROM SIX SUBJECTS

FOR THE LOOKS CONDITION OF EXPERIMENT I

TRIAL	SERIES 1		SERIES 2		SERIES 3		MEAN	
	Yes	No	Yes	No	Yes	No	Yes	No
27	545	563	469	489	459	487	491	513
44	488	510	475	462	422	483	462	485
26	586	574	464	490	407	455	486	506
34	555	545	424	473	455	456	478	491
3	434	478	408	416	373	417	405	437
41	542	606	472	491	468	491	494	529
45	557	504	487	526	465	495	503	508
9	393	465	395	455	375	563	388	461
11	395	460	480	389	348	393	374	414
38	395	536	462	485	492	485	500	502
7	545	421	350	424	363	381	366	409
25	385	504	451	538	458	532	463	525
54	481	752	581	693	567	672	600	706
49	653	588	545	551	531	500	560	546
30	604	593	476	489	507	440	509	507
43	544	583	528	583	509	527	552	564
5	417	462	412	447	402	453	410	454
47	746	597	614	583	497	555	619	578
22	508	545	473	449	371	452	451	582
29	464	482	472	470	430	484	455	479
36	532	522	513	496	434	450	493	489
6	419	453	409	424	400	461	409	446
42	529	577	502	525	507	494	513	532
46	633	613	616	594	571	534	607	580
16	518	510	464	516	414	482	465	503
55	553	576	553	518	458	461	521	518
21	496	501	452	529	405	492	451	507
2	415	562	437	478	364	437	405	492
31	462	496	461	475	455	446	459	472
39	551	613	488	568	481	515	507	565
48	706	631	590	586	536	552	611	570
8	435	489	436	435	366	402	412	442
10	412	443	360	414	352	383	375	413
28	459	498	461	498	410	461	443	486
1	396	496	353	407	374	415	374	439
53	529	644	648	661	571	596	599	634
52	655	625	654	612	569	595	626	611
24	534	505	487	515	435	454	485	491
40	582	628	643	550	501	534	575	571
4	381	461	443	469	393	443	406	458
50	488	522	484	505	476	583	486	537
18	492	507	402	465	392	444	429	472
35	554	544	540	559	500	487	531	530
33	481	510	396	471	458	474	445	485
13	463	530	484	504	413	466	453	500
51	625	619	498	581	548	594	557	598
12	444	524	371	445	367	415	334	461
32	463	469	414	469	467	465	448	468

APPENDIX C

TRIAL BY TRIAL REACTION TIME MEANS FROM SIX SUBJECTS

FOR THE MEANS CONDITION OF EXPERIMENT I

TRIAL	SERIES 1		SERIES 2		SERIES 3		MEAN	
	Yes	No	Yes	No	Yes	No	Yes	No
27	540	635	518	582	557	550	538	589
44	542	536	511	536	495	541	516	538
26	561	563	488	512	469	518	506	531
34	489	511	456	474	418	471	454	485
3	451	509	450	451	381	482	427	481
41	745	654	575	559	527	586	616	600
45	562	603	521	562	527	529	537	565
9	433	512	402	446	363	461	399	473
11	445	505	434	472	419	468	433	482
38	629	606	523	498	554	511	569	438
7	519	526	453	466	430	464	467	485
25	513	562	486	534	459	529	486	545
54	500	521	464	486	480	476	481	494
49	555	553	498	547	560	537	538	546
30	539	632	510	513	471	505	507	550
43	710	668	634	616	579	611	641	632
5	473	522	472	476	395	484	447	494
47	730	858	764	830	707	754	734	814
22	489	504	415	464	471	478	458	482
29	549	533	470	501	487	429	502	488
36	521	551	494	488	543	531	519	523
6	419	475	375	449	447	448	414	457
42	568	559	506	502	534	517	536	526
46	560	539	478	534	527	523	522	532
16	420	491	449	480	431	473	433	481
55	537	520	517	479	503	471	519	490
21	518	466	449	477	425	459	464	467
2	442	494	420	508	453	513	438	505
31	471	550	449	454	429	482	450	495
39	566	597	535	582	538	572	546	584
48	599	673	587	598	548	582	578	618
8	519	509	461	476	444	504	475	496
10	450	496	460	446	399	470	436	471
28	525	551	469	489	477	517	490	519
1	437	490	428	487	428	452	431	476
53	626	580	609	557	532	570	589	569
52	614	674	589	566	573	586	592	609
24	520	511	476	454	465	484	487	483
40	569	676	503	535	473	570	515	594
4	505	512	499	467	499	489	501	489
50	504	499	468	506	501	460	491	488
18	459	502	416	468	439	458	438	476
35	621	665	511	550	517	554	550	590
33	480	500	390	483	434	458	435	480
13	465	539	411	483	.412	469	429	497
51	690	669	661	613	605	640	652	641
12	432	533	417	504	357	454	402	497
32	450	478	393	432	435	442	426	451

APPENDIX D

TRIAL BY TRIAL REACTION TIME MEANS FROM SIX SUBJECTS

FOR THE SOUNDS CONDITION OF EXPERIMENT I

TRIAL	SERIES 1		SERIES 2		SERIES 3		MEAN	
	Yes	No	Yes	No	Yes	No	Yes	No
27	447	535	414	482	409	454	423	490
44	488	506	445	484	439	400	457	463
26	467	512	435	441	411	461	438	471
34	433	480	419	455	418	444	423	460
3	439	511	388	434	424	420	417	455
41	472	490	427	486	419	464	439	480
45	500	532	502	462	468	451	490	482
9	418	490	385	449	429	447	411	462
11	430	452	440	423	379	397	416	424
38	507	518	462	480	441	459	470	486
7	384	465	408	463	391	434	394	454
25	466	553	442	455	432	458	447	489
54	488	491	448	522	419	467	452	493
49	513	548	505	529	508	497	509	525
30	487	541	479	499	416	480	461	507
43	512	512	418	467	472	494	467	491
5	471	457	413	427	392	417	425	434
47	481	524	476	476	476	472	478	491
22	489	515	409	472	432	443	443	477
29	522	522	436	466	468	467	475	485
36	525	613	498	555	481	520	501	563
6	465	523	403	456	428	440	432	473
42	532	519	448	491	470	491	483	500
46	539	555	453	467	449	482	480	501
16	440	465	375	457	378	417	398	446
55	441	485	435	473	450	478	442	479
21	444	517	416	469	405	449	422	478
2	476	494	395	490	404	464	425	483
31	424	484	430	478	414	455	423	472
39	481	498	452	468	472	468	468	478
48	468	544	479	531	496	478	481	518
8	445	468	413	428	393	441	417	446
10	475	486	412	445	404	436	414	456
28	494	489	462	483	437	469	464	480
1	412	480	389	469	437	435	413	461
53	461	503	423	481	397	440	427	475
52	505	525	518	488	432	516	485	510
24	438	483	433	437	391	476	421	465
40	470	520	445	455	457	523	457	499
4	456	497	396	433	391	454	414	461
50	479	485	456	469	552	459	496	471
18	418	409	431	483	376	447	408	446
35	472	544	457	500	412	486	447	510
33	389	431	388	447	375	427	384	435
13	458	486	409	444	410	503	426	478
51	488	533	439	471	503	472	477	492
12	461	521	430	458	411	497	434	492
32	409	486	380	430	403	441	397	452

APPENDIX E

TRIAL CONTENTS FOR SOUNDS 1:3 OF EXPERIMENT II

Trial Number	Contents	Auditory Units	Trial Number	Contents	Auditory Units
1	WON	1	25	1 2 8	3
2	2	1	26	WON TOO ATE	3
3	4	1	27	1 TWO 4	3
4	EIGHT	1	28	WON TWO 4	3
5	1 ONE	1	29	1 ONE 2 8	3
6	TWO TOO	1	30	TWO 4 8 EIGHT	3
7	FOUR FOR	1	31	1 ONE TOO FOR	3
8	8 ATE	1	32	ONE WON 2 EIGHT	3
9	1 ONE WON	1	33	TWO FOUR EIGHT TOO FOR ATE	3
10	2 TWO TOO	1	34	1 2 4 ONE TWO FOUR	3
11	4 FOUR FOR	1	35	1 WON TWO TOO 8 EIGHT	3
12	8 EIGHT ATE	1	36	TWO 4 FOUR FOR 8 EIGHT	3
13	4 8	2			
14	FOR ATE	2	37	1 2 4 8	4
15	ONE 2	2	38	WON TOO FOR ATE	4
16	WON FOUR	2	39	1 TOO 4 ATE	4
17	ONE WON EIGHT ATE	2	40	WON TOO 4 EIGHT	4
18	FOUR FOR EIGHT ATE	2	41	ONE TWO FOUR FOR EIGHT	4
			42	1 ONE 2 FOUR 8	4
19	2 TOO 4 FOUR	2			
20	2 TOO FOUR FOR	2	43	WON 2 FOR EIGHT ATE	4
21	1 ONE WON 2 TWO TOO	2	44	1 ONE TWO FOR 8	4
22	1 ONE WON 8 EIGHT ATE	2	45	1 2 4 8 ONE TWO	4
23	2 TWO TOO 4 FOUR FOR	2	46	WON TOO FOUR FOR EIGHT ATE	4
24	4 FOUR FOR 8 EIGHT ATE	2			
			47	ONE WON 2 TOO FOUR ATE	4
			48	WON TWO 4 FOUR EIGHT ATE	4

Practice Trials

51	TWO ATE 1 TOO	3
52	8	1
53	ONE 2 FOUR FOR 8 ATE	4
54	8 4	2
55	ONE TWO EIGHT	3
56	4 WON TOO FOR ATE	4
57	TWO FOUR ATE	3
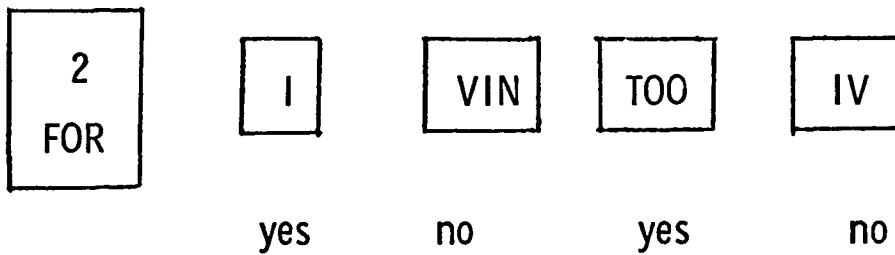
APPENDIX F

ILLUSTRATED INSTRUCTIONS FOR EXPERIMENT II

"The   Same   Set"

| 1 | 2 | 4 | (arabic)

| I | II | IV | (roman)

| ONE | TWO | FOUR | (english #)

| WON | TOO | FOR | (english words)

Example:

| 2
| FOR

| I | | VIN | | TOO | | IV |

yes   no   yes   no

APPENDIX G

MEAN REACTION TIMES LISTED BY SEQUENTIAL CATEGORIES & SET SIZE

Condition	Exp.	Res.	Cat.	1	2	3	4	5	6
Looks 1:1	I	Yes	ID	372	429	444	476		
			EQ	–	449	482	556		
			NR	410	476	517	568		
		No	ID	387	477	541	532		
			EQ	460	508	544	575		
			NR	428	488	506	570		
	III	Yes	ID	413	451	489	518	531	610
			EQ	–	477	578	567	638	683
			NR	433	522	544	586	599	654
		No	ID	437	624	580	589	609	552
			EQ	515	530	623	624	694	728
			NR	475	583	607	604	652	732
Sounds 1:3	I	Yes	ID	380	415	460	429		
			EQ	415	449	495	494		
			NR	422	459	475	489		
		No	ID	416	464	452	533		
			EQ	462	487	513	484		
			NR	454	482	499	490		
	II	Yes	ID	398	395	423	431		
			EQ	429	459	477	467		
			NR	452	487	492	500		
		No	ID	507	483	499	487		
			EQ	527	550	556	546		
			NR	514	527	527	517		
Sounds 1:1	II	Yes	ID	369	427	424	425	466	440
			EQ	–	446	474	501	555	570
			NR	409	487	502	516	532	551
		No	ID	454	546	561	550	554	591
			EQ	483	562	588	566	584	622
			NR	464	529	544	563	580	592
	EXT.	Yes	ID	325	364	373	389	387	427
			EQ	–	388	413	421	455	470
			NR	331	385	418	441	467	484
		No	ID	387	496	491	468	470	494
			EQ	398	454	482	480	497	527
			NR	384	439	463	468	501	509