Presentation Conditions and Psychological Context: An Investigation of List and Position Tags

Kathleen Ann Carlson
Loyola University Chicago

Recommended Citation
Master's Theses. 2844.
https://ecommons.luc.edu/luc_theses/2844
PRESENTATION CONDITIONS AND PSYCHOLOGICAL CONTEXT:

AN INVESTIGATION OF LIST AND POSITION TAGS

by

Kathleen A. Carlson

A Thesis Submitted to the Faculty of the Graduate School of Loyola University of Chicago in Partial Fulfillment of the Requirements for the Degree of Master of Arts

February

1976
ACKNOWLEDGMENTS

The author gratefully acknowledges the assistance and encouragement provided by Dr. Eugene B. Zechmeister at all stages of this project. Thanks are also extended to Dr. Frank Slaymaker and Dr. David DePalma for their helpful suggestions during the preparation of this manuscript.
Kathleen Ann Carlson was born on June 28, 1948 in Chicago, Illinois. She is the only daughter and youngest child of Earl and Margaret Carlson's three children. She graduated from Alvernia High School, a parochial school for women run by the School Sisters of St. Francis, in 1966. She then attended Loyola University of Chicago, where she received a Bachelor of Science degree in Psychology (magna cum laude) in June, 1970.

After graduating from college, Kathleen served on a vice-presidential staff of Illinois Bell Telephone Company as an assistant staff supervisor. In that position, she was responsible for collecting and analyzing accident data for more than 10,000 Chicago area employees. Her research experience with Illinois Bell motivated Kathleen to pursue a doctor's degree in Experimental Psychology.

In September, 1972, Kathleen entered the graduate program at Loyola, where she has concentrated her interests in human learning and memory, and visual information processing. Upon completion of her graduate work, Kathleen will seek an academic position in which it will be possible to teach and conduct research.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acknowledgements</td>
<td>ii</td>
</tr>
<tr>
<td>Life</td>
<td>iii</td>
</tr>
<tr>
<td>List of Tables</td>
<td>v</td>
</tr>
<tr>
<td>List of Figures</td>
<td>vi</td>
</tr>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Two Alternative Hypotheses for Recency</td>
<td>5</td>
</tr>
<tr>
<td>Strength Hypothesis</td>
<td>5</td>
</tr>
<tr>
<td>Critique of Strength Hypothesis</td>
<td>11</td>
</tr>
<tr>
<td>Context Hypothesis</td>
<td>17</td>
</tr>
<tr>
<td>Evaluation of Context Hypothesis</td>
<td>21</td>
</tr>
<tr>
<td>On the Nature of Contextual Tags</td>
<td>23</td>
</tr>
<tr>
<td>Method</td>
<td>31</td>
</tr>
<tr>
<td>Results</td>
<td>37</td>
</tr>
<tr>
<td>List Accuracy</td>
<td>41</td>
</tr>
<tr>
<td>Position Accuracy</td>
<td>43</td>
</tr>
<tr>
<td>List-Plus Accuracy</td>
<td>45</td>
</tr>
<tr>
<td>Effects of Duration</td>
<td>48</td>
</tr>
<tr>
<td>Mean Placement of Items</td>
<td>52</td>
</tr>
<tr>
<td>Discussion</td>
<td>55</td>
</tr>
<tr>
<td>Conclusions</td>
<td>64</td>
</tr>
<tr>
<td>References</td>
<td>66</td>
</tr>
</tbody>
</table>
List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Raw Number Correct by Groups</td>
<td>39</td>
</tr>
<tr>
<td>2</td>
<td>Response Frequency of Categories</td>
<td>40</td>
</tr>
<tr>
<td>3</td>
<td>Position Data into Separate Components</td>
<td>46</td>
</tr>
<tr>
<td>4</td>
<td>Mean Confidence for All Subjects</td>
<td>49</td>
</tr>
<tr>
<td>5</td>
<td>Accuracy for Items of Each Duration</td>
<td>51</td>
</tr>
<tr>
<td>6</td>
<td>Mean Placement for Items of Each Duration</td>
<td>54</td>
</tr>
</tbody>
</table>
List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>List Accuracy by Groups</td>
<td>42</td>
</tr>
<tr>
<td>2</td>
<td>Position Accuracy by Groups</td>
<td>45</td>
</tr>
<tr>
<td>3a</td>
<td>List-Plus Accuracy for Short Groups</td>
<td>47a</td>
</tr>
<tr>
<td>3b</td>
<td>List-Plus Accuracy for Long Groups</td>
<td>47b</td>
</tr>
</tbody>
</table>
INTRODUCTION

Contemporary memory models have conceptualized the long-term storage of memories as a collection of attributes that serve to discriminate one memory from another (e.g., Underwood, 1969). When we retrieve a particular memory, we may recall not only the occurrence of some past event, but also some idea as to when the event happened (temporal attribute), where it occurred (spatial attribute), as well as verbal and nonverbal associations of the memory (associative attributes). According to this perspective, a memory cannot be retrieved without its attributes, since the attributes are the only means by which one memory is identified as being separate from all other memories.

The present paper deals mainly with the temporal attribute of memory. The reasons for this interest are twofold. First, there is sound empirical evidence to suggest that a temporal attribute plays a fundamentally important role in memory retrieval. Underwood (1969), for example, reported that by increasing the time interval between paired-associate lists, inter-list interference was drastically reduced (i.e., the increased temporal separation resulted in discriminable temporal attributes associated with each list). Likewise, Light and Schurr (1973) demonstrated that providing subjects with temporal cues in a recognition memory task (by having test items presented in the same temporal order in which they were...
studied) significantly improved their performance.

A second reason for investigating the temporal attribute stems from the fact that it has received relatively little theoretical or empirical attention. Tulving and Madigan (1970), after a thorough review of the pertinent literature, concluded that psychologists have been somewhat reluctant to incorporate the temporal attribute within any precise theoretical framework. Thus, virtually nothing is known about how temporal information is stored in memory, or how it operates as an effective discriminative cue.

Beginnings of Empirical Research

A test paradigm appropriate for the empirical investigation of the temporal attribute, usually referred to as a "judgment of recency task," was introduced by Yntema and Trask (1963). In this study, a long list of words printed singly on index cards was inspected by subjects, each at his own pace. Interspersed among study words were test cards, each one containing two words followed by a question mark. When a test card appeared, subjects were asked to choose the word that occupied the more recent position in the preceding inspection series. Test pairs were constructed so that either one or both of the test items actually appeared in the inspection series. In cases where only one member of the test pair appeared in the series, subjects were instructed to consider "new" test items
as having occurred "very many words ago."

By the method outlined above, Yntema and Trask wished to assess two major components of recency discrimination: (a) separation interval, or the number of items that intervened between the inspection of each word comprising a test pair; and (b) lag, the number of items that intervened between the time of inspection and the time of test for the more recent member of a test pair. The range of lag and separation values manipulated in the Yntema and Trask study was considerable, with the lowest value at 4 items, and the highest value reaching 136 items. It was found that recency judgments, which were far better than chance at all lag and separation values, became more accurate as the separation interval increased, but less accurate as the lag interval increased.

On the basis of their results, Yntema and Trask concluded that the temporal attribute of memory is represented by hypothetical "time tags" that a subject examines to determine the relative recencies of past events. It was also suggested that time, or something correlated with time, comprises the time tags in memory. Thus, the observed effect for separation interval was explained by the fact that values for this interval are perfectly correlated with time. That is, the larger the interval between members of a test pair, the larger the time difference between their occurrence, and presumably, the more discriminable their time tags would be. Likewise, as the time interval between the time of study and
the time of test (lag) increases, the discriminability of the time tags associated with test items must decline; and this was suggested to account for the observed lag effect.

The concept of time tags, as presented in Yntema and Trask (1963), has been criticized by a number of cognitive psychologists. Murdock (1974), for example, argues that it is merely a descriptive label, pointing to the fact that people can discriminate memories on the basis of recency. Also, Estes (1972) concluded that the notion of time tags does not address itself to the problem of explaining how temporal information is represented in memory; and it is, therefore, atheoretical speculation.

Despite these criticisms, however, the concept of time tags has stimulated further research into the problem of recency. Moreover, because of the theoretical weaknesses of the time tags hypothesis, psychologists from various theoretical backgrounds have sought to find a better, alternative explanation. Among the alternatives offered, two major hypotheses (the strength hypothesis and the context hypothesis) have emerged as rival explanations for recency phenomena. Each of these hypotheses will now be reviewed in detail.
TWO ALTERNATIVE HYPOTHESES FOR RECENTY

The Strength Hypothesis

Strength theory represents a comprehensive model of memory that has been formally presented elsewhere (Anderson & Bower, 1972; Kintsch, 1970; Murdock, 1974; Wickelgren, 1970). However, there are three assumptions of this model that are pertinent to understanding the strength hypothesis for recency. First, it is assumed that a memory is stored as a single, unidimensional "trace." The exact nature of a trace is not specifically defined, other than the fact that it is thought to bear some physiological relationship to central neural circuits.

Second, it is thought that activation (i.e., retrieval) of a memory trace is accompanied by an implicit and direct estimate of its strength. The concept of trace strength has been rather loosely defined, but it is often equated with the variable of familiarity; the stronger the memory trace, the more familiar the memory represented by the trace appears. Trace strength is also presumed to have a direct influence upon the ease with which items are retrieved from memory. Strength theorists generally infer that the easier it is to recall or recognize an item, the stronger its trace strength must be.

The third and most important assumption of the model is that trace strength diminishes, in either discrete or continuous fashion, with the passage of time. The strength
hypothesis for recency grew directly from this assumption, and it was summarized succinctly by Konorski (1961):

We have a strong inclination to believe that the "sense of time" of men and other animals i.e., the sense of varying durations of time which have elapsed since a definitive event, is based on nothing else than the strength of traces left by this event at various moments after its cessation. The weaker these traces the more remote in time the given event seems to be. (p. 122)

Thus, according to the strength hypothesis, recency judgments are mediated by the single variable of trace strength, and are considered to be the by-products of the trace decay process.

The strength hypothesis, as outlined above, has generated three major predictions for recency phenomena. The first prediction specifies that variables such as frequency and repetition, which are thought to increase the strength of memory traces, should also enhance the apparent recency of memories represented by such traces. To test this prediction, Peterson (1967) devised a recency task in which a long list of verbal stimuli were presented at a fixed rate. At various times throughout the inspection list, subjects were shown a test word that had been seen in the preceding series, and were required to estimate lag, or the number of items that had occurred since the test word was last seen. The major independent variable in this study was stimulus frequency, the number of times test words were repeated in the inspection series before the time of test. Half of the items were seen only once, whereas the other half were seen twice before appearing as a test item. Lag intervals ranged
from 2 to 8 items, and the first and second presentations of repeated items were separated by four intervening words. If the strength hypothesis was correct, repeated items should have stronger traces than nonrepeated ones, and, therefore, should appear more recent than nonrepeated items.

Contrary to this prediction, however, no significant differences between repeated and nonrepeated items were observed. To interpret this finding, Peterson modified the strength hypothesis to suggest that items repeated in a spaced fashion (i.e., repetitions separated by several intervening items) result in two separate traces with strengths no greater than singly presented items. Since the repeated items in that study were presented in a spaced manner, their lag estimates did not differ from the once-presented items for the simple reason that their trace strengths were not appreciably different.

It was further suggested that items presented in a massed condition (i.e., repetitions in successive order) would result in a single, strengthened trace, relative to the traces of spaced items. Thus, it was predicted that massed items should appear more recent than spaced items at the same lag interval. In a later experiment (Peterson, Johnson, & Coatney, 1969) this prediction was tested, and it was found that Peterson's hypothesis was confirmed. The lag estimates for massed items were significantly smaller than those for spaced items.
When reviewing the Peterson experiments, one should note that the test paradigm differed somewhat from the task originally introduced by Yntema and Trask (1963). Peterson observed absolute judgments of recency for individual stimuli, a task that does not entail the discrimination of recency for two different test items (i.e., a comparative task). Lockhart (1968) was interested in seeing whether the strength hypothesis, which was tested by Peterson in an absolute task, also accounts for recency data observed in a comparative task. To do this, Lockhart tested whether performance by one group of subjects in an absolute task could predict the performance by another group in a comparative task. If comparative judgments are made on the basis of lag estimates for each member of a test pair, as the strength hypothesis implies, then a probability distribution generated from an absolute task should predict the data observed in a comparative one. Lockhart's data confirmed the strength model's predictions, and it was concluded that the processes envisaged by the strength hypothesis for absolute tasks also characterizes the processes involved in comparative ones.

A second major prediction of the strength hypothesis pertains directly to comparative judgments of recency. Namely, when two items of unequal strength are presented in a test pair, recency discrimination will be more accurate when the item with the stronger trace is actually the more recent member of the test pair. Likewise, if the stronger trace
precedes the weaker one in time, recency discrimination will be impaired. To test this prediction, Morton (1968) observed comparative recency judgments under three experimental conditions: (a) when the more recent item was repeated earlier in the inspection series; (b) when the less recent item was repeated; and (c) when neither item was repeated. According to the strength hypothesis, repetition of the less recent member should increase its trace strength relative to the more recent one; hence, recency discrimination should be impaired under that condition.

Test stimuli consisted of numerical digits presented auditorially in a series of 14 other digits. Digits that were repeated had only one item separating their first and second presentations. In agreement with the strength hypothesis, Morton found that judgments were most accurate when the more recent item was repeated, and least accurate when the less recent item was repeated.

Further evidence in support of this prediction has been found in cases where test stimuli were pictoral, rather than verbal in nature. It is a well-known fact that pictures are easier to remember than words (e.g., Pavio, 1971). A strength theorist interprets this to mean that the trace strength of pictures declines less rapidly than that of verbal stimuli. As applied to recency judgments, the strength hypothesis predicts that pictures should appear relatively more recent than words at constant lag intervals. Fozard
and Weinert (1972) compared absolute judgments of recency for pictures and words, and found that at relatively short lags (13 to 14 items), pictures appeared more recent than words. At longer lags, however, the differences between pictures and words disappeared. Furthermore, Fozard (1970) observed comparative recency judgments when test pairs consisted of both a word and a picture. It was found that recency discrimination was impaired when the more recent member was a word. This latter finding is in agreement with the Morton (1968) study, since it is assumed that the trace strength of a picture is stronger than that of a word.

A third major prediction of the strength model involves the relationship between trace decay and decision criteria that subjects use to transform values of strength into estimates of lag. Hinrichs (1970), by applying the principles of statistical decision theory, offered a description of this transformation process. He assumed that values of trace strength are normally distributed about mean values of lag. It was further assumed that subjects implicitly establish criterion values for strength that demarcate the range of values associated with each lag estimate. Thus, if the trace strength for a given stimulus falls short of, or exceeds a particular criterion value, the lag estimate for that stimulus is determined by the lag value associated with the range of strength values falling above or below that criterion.
According to the strength hypothesis, trace strength declines as a function of time, and is independent of any criteria subjects may use to mediate recency judgments. To test this assumption, Hinrichs (1970) deceived subjects into believing that the maximum lag interval they would experience in an absolute task was either 6, 9, or 12 items in length. Actually, the maximum possible lag was always 9 items. By this deception, it was felt that the decision criteria were varied independently of the decay rate of trace strength. It was therefore predicted that the rate of trace decay, as evidenced by the rate with which mean estimates of lag declined with time, should be identical in all three deception conditions. The results fit the strength model's predictions, and Hinrichs concluded that trace strength declines independently of decision criteria.

Critique of Strength Hypothesis

Although the strength model has enjoyed some empirical verification, cogent arguments against its acceptance have been waged. Most adversaries have based their argument on the basic criticism that the strength hypothesis is far too simplistic. Its simplicity stems from the fact that it relies exclusively upon the single variable of trace strength to explain recency data. In effect, the strength hypothesis denies the utility of all other attributes of memory that a person may use to determine an event's occurrence in time.
Such extreme simplicity often renders the strength hypothesis incapable of resolving some fundamental logical and empirical problems.

Some of the logical problems attendant to the strength hypothesis were noted by Underwood (1969). One has only to reflect upon his own personal experiences to realize that some older memories appear to be much more vivid or "stronger" than some more recent ones. Yet, we are still capable of remembering the temporal order from older to newer memories. If recency discrimination were based solely on trace strength, considerable disorientation in time would result. Therefore, the strength model is at a loss to explain how older, more vivid memories are kept in their proper temporal position.

One of the most poignant statements of criticism against the strength hypothesis was presented by Tulving and Madigan (1970):

In our opinion, the strength hypothesis is a product of desperation. It is entirely possible that in the absence of any other information the subject may correctly or incorrectly reason that of the two items, the one appearing more familiar may look so because because it appeared more recently, but this does not mean that the subject has no access to more direct information about the temporal code of an item in many other situations. (p. 463)

In effect, Tulving and Madigan suggest that a simple, unidimensional variable such as trace strength mediates recency judgments only when there are no other attributes associated with items stored in memory. Thus, it would be wrong to conclude that the sense of recency is due solely to the strength of traces representing stored memories.
In support of Tulving and Madigan's conclusion, it can be argued that the variable of trace strength accounts for recency data in only a few experimental settings. Moreover, many of the strength theorists' conclusions regarding the efficacy of the strength variable must be restricted to situations where only short lag intervals are being observed. A close examination of much of the data cited in support of the strength hypothesis reveals that at longer lag intervals, subjects were guessing their recency judgments. Hinrichs and Bushke (1968), for example, reported that the accuracy of lag estimates (in an absolute type of task) declined to chance performance at lag intervals greater than 8 items. Likewise, in the Morton (1968) study described earlier, subjects were making comparative judgments of recency at no better than chance probability in two of the three experimental conditions. As Underwood (1969) indicated, if subjects are merely guessing their decisions, it is erroneous to conclude that any variable, including trace strength, is responsible for the results.

A second limitation to the conclusions of the strength theorists stems from the fact that their predictions have been upheld in only one type of experimental task, the continuous or probe type of task. Continuous tasks are characterized by the fact that there is no temporal separation between the study and test phases. Recency judgments are made throughout the presentation of a long series of stim-
uli. Bower (1972) has suggested that continuous tasks are "impoverished" in the sense that subjects are afforded little opportunity to make recency judgments on factors other than familiarity or strength. This impoverishment is exemplified by the fact that there are no distinctive spatial or visual cues that would help discriminate one stimulus from another. Furthermore, the forced presentation rate, coupled with the temporal contiguity between the study and test phases, do not allow subjects the opportunity to cognitively organize or rehearse the tested items. Given such circumstances, it is not surprising that subjects' performance rarely exceeds chance level at longer lags, or that the variable of trace strength stands out as the most likely explanation for the sense of recency.

Continuous types of tasks, however, are by no means the only methods available for the investigation of recency phenomena. There is another class of tasks that can be termed "discrete" in that recency judgments are delayed until the completion of the entire inspection series. Discrete tasks differ from continuous ones in two important ways: (a) lag intervals in discrete tasks can vary over a much wider range of values, from a few items to a few days; and (b) in discrete tasks, there is a clear temporal separation between the study and test phases of the experiment. Because of these characteristics, discrete tasks bear a stronger resemblance to real life in that recency judgments are made from
a retrospective vantage point from which each item is considered as a member of an entire series of past events.

There is a growing body of evidence to show that predictions of the strength hypothesis are not upheld in discrete types of tasks. Underwood (1969) cited several unpublished studies in which absolute judgments of recency were measured by having subjects identify the temporal position of stimuli with numbers corresponding to each item's position in a presentation list. Test words differed from each other in terms of their study time, and it was found that the longer a word had been studied, the easier it was recalled on a later test. Strength theory interprets this to mean that items studied for a longer time have stronger traces than items studied for shorter times. Accordingly, stronger items should appear more recent (i.e., assigned to more recent temporal positions) than weaker ones. Contrary to the strength hypothesis' prediction, however, position judgments did not vary systematically with study time. Items studied for a longer time were not assigned to more recent serial positions. Underwood therefore concluded that there is no correlation between trace strength, as manipulated by study time, and judgments of recency.

Another failure of the strength hypothesis was noted by Winograd (1968), who indirectly measured recency judgments in a list discrimination task. Subjects were shown two word lists such that List 1 was presented in its entirety
a variable number of times before List 2 was presented. In some cases, List 1 was repeated 9 times more frequently than List 2 was presented. After List 2 was presented for the last time, subjects were asked to recall the presentation list in which test words appeared. Strength theory assumes that the trace strength of list items increases with the frequency of repetition. Thus, in cases where the frequency of List 1 was much greater than that for List 2, list discrimination should be impaired. This expectation is identical to the second major prediction of the strength hypothesis outlined earlier (cf. Fozard, 1970; Morton, 1968). But, contrary to the strength hypothesis, Winograd found that the accuracy of list discrimination increased, rather than decreased, as the frequency of List 1 increased.

Further evidence against the strength hypothesis has also been reported by Hintzman and Block (1971). In this study, subjects were shown a fairly long list of words, and were later asked to recall the temporal position of test words randomly drawn from the presentation list. A horizontal line divided into ten equal segments was printed at the top of each answer sheet, and position judgments were made by assigning a number from 1 to 10 that corresponded to the tenth of the list in which each test item appeared. It was found that position judgments were far better for words from the beginning of the list than those from the middle or latter portions.
This primacy effect observed by Hintzman is difficult for the strength hypothesis to explain. If recency is represented in memory only by the strength of an exponentially decaying trace (cf. Hinrichs, 1970), it is not clear why the temporal positions of less recent items are better discriminated than the temporal positions of more recent items. A strength theorist could argue that items from the beginning of a list have stronger traces (i.e., because of less interference) than items from other portions. But, in that case, the strength model would also have to predict that beginning items be assigned to more recent serial positions. Clearly, the data do not support any explanation offered by the strength hypothesis, and an alternative explanation is needed to account for Hintzman's primacy effect.

The Context Hypothesis

The context model of memory differs from the strength model in several important ways. First, the context model assumes that a memory is multidimensional i.e., comprised of many discrete elements. The elements comprising a memory have been collectively termed the "stimulus-as-coded" (s-a-c), and are thought to represent stimulus features encoded into memory (Bower, 1972). Consequently, the representation of an item in memory depends upon the manner in which it was encoded. Further, encoding is assumed to depend upon a subject's attentional set.
The context model also assumes that associations are immediately and implicitly formed between a s-a-c and a multitude of other s-a-c's that correspond to contextual stimuli present at the time of encoding. Therefore, when an item is retrieved from memory, contextual elements associated with the item are also retrieved. Recognition memory is seen to depend upon the retrieval of contextual associates. To recognize that an item was seen before, one must implicitly realize, through contextual retrieval, that the item was seen in a previous context.

The context model, unlike the strength model, emphasizes the associative nature of memory to explain the sense of recency. As mentioned earlier, strength theory relies upon the single variable of trace strength to explain recency judgments. The context model, however, proposes that the sense of recency is mediated by the retrieval of a multitude of elements associated with an item in memory.

Proponents of the context model have suggested at least two ways in which a person utilizes contextual elements to mediate judgments of recency. The first method, which can be termed a "statistical scanning process," was proposed by Bower (1972) as an explanation for recency judgments measured in continuous types of tasks. When an item is presented for test, the subject supposedly is able to scan the set of retrieved elements and estimate the proportion that are "tagged" (i.e., associated with) to the item's previous experimental
context. According to Bower, the greater the proportion of tagged elements found, the more recent the item appears. It was further assumed that with longer time intervals between the item's presentation in the inspection series and test (lag), the elements tagged to the previous context would have a lower probability of being retrieved with the test item. Thus, for absolute judgments of recency, the greater the lag interval, the less recent the item appears.

As applied to comparative judgments of recency, the context hypothesis proposes that the test item having the greater proportion of tagged elements is the one that will appear more recent. As the separation interval between items to be compared increases, the greater the likelihood of differences between the proportion of tagged elements retrieved with each item. Thus, the greater the separation interval between items to be compared, the more discriminable (in terms of recency) the test items become. The context model is therefore in agreement with the finding that recency discrimination becomes more accurate as the separation interval increases (cf. Yntema and Trask, 1963).

A second way in which subjects use contextual elements is through contextual "tags" or "markers." According to Anderson and Bower (1972), contextual tags are hypothetical labels that summarize and define a multitude of elements appearing within a discrete interval of time. The notion of contextual tags has been used to explain judgments of serial
position (e.g., Hintzman and Block, 1971). It is thought that "list tags" and "position tags" help a subject pinpoint the serial position in which an item appeared. As Anderson and Bower (1972) explained:

The point of introducing associations to context is to provide a means of keeping track of the occasions in which particular words appeared. This would be difficult to implement on the basis of direct associations between the word and the contextual elements. How would the subject know which contextual elements belonged to which list? (p. 104)

Presumably, the greater the differences between contexts encoded with list items, the more likely it is that subjects will tag them differentially. As a test of this prediction, Brown (1973b) asked school children to make comparative judgments of recency for pairs of picture stimuli presented under varying conditions of context. In the Visual condition, each group of 8 items was presented in different spatial locations representing various points along a child's journey to school (e.g., house, garden, street etc.). In the List condition, pictures were shown in blocks of 8 items, but without extra spatial cues. Finally, in the Blank condition, all pictures were presented in an uninterrupted series.

The results were in total agreement with the predictions of the context model. For older children, recency discrimination was better in the Visual condition (77%) than in either the List (69%) or Blank (63%) conditions. Furthermore, judgments were most accurate when the pictures comprising test pairs came from different contexts. In the
Visual condition, for example, performance was 85% accurate when items came from different contexts; whereas, judgments were only 52% accurate when no such contextual differences existed. These findings were later replicated when context was varied by having different colored backgrounds for blocks of study items (Brown, Campione, and Gilliard, 1974).

**Evaluation of the Context Hypothesis**

The context model, unlike the strength model, can provide explanations for recency data from a variety of experimental tasks. The theoretical superiority of the context model in this respect can be demonstrated by reviewing two studies mentioned earlier. First, in the Winograd (1968) study, the strength model would have erroneously predicted that list discrimination would decrease as the frequency of List 1 increased. The context model, on the other hand, could argue that the probability of retrieving the List 1 tag would increase with the frequency of its repetition. Thus, the context model would agree with the Winograd data showing improved list discrimination with List 1 repetitions.

Second, Underwood (1969) observed that the strength model incorrectly predicted that "strong" items presented at the beginning of the list would appear more recent than "weaker" items from the end of the list. According to the context model, however, the strength or vividness of an item has no bearing on its apparent recency other than to
the degree to which vividness influences the formation of
discriminable list markers. Thus, the context model would
not contradict our personal observations that temporal or-
ientation is preserved, even though past events differ in
terms of their vividness or familiarity.

In light of the above discussion, the context hypo-
thesis appears to be a desirable alternative to the overly-
simplistic notions of strength theory. On the negative
side, however, the context model has suffered one major
criticism. Namely, the term "context" has been treated in
much too general a fashion. "Context," as it was originally
described by Anderson and Bower (1972), encompasses an almost
unlimited variety of elements, ranging from internal physio-
logical cues, to the physical parameters of the external
environment. According to Brown (1973a), the context model
has thus far been unable to specify which, among the multi-
tude of contextual elements, are most influential in the for-
mation of list or position tags. As Wells (1974) pointed out:

The nature of contextual tags needs careful specification
to avoid the emptiness of the statement that one re-
members context because one remembers contextual infor-
mation. (p. 390)
ON THE NATURE OF CONTEXTUAL TAGS

Hintzman, Block, and Summers (1973), in a study of serial list and position judgments, attempted to clarify the nature of contextual tags, at least as they operate in studies of position knowledge. Subjects were shown four word lists separated by a 2.5 minute recognition test. After the last list was shown, serial position judgments were measured by having subjects indicate both the list and within-list position in which test items appeared.

There were three major findings in this study. First, there were primacy and recency effects. The positions of items from the beginning and end of the inspection series were better discriminated than those from other portions of the series. The recency effect, however, was found to be transitory in that prolonging the time interval between the study and test phases of the experiment attenuated it. Second, errors of list discrimination tended to fall in lists that were in closest temporal proximity to the target list. Third, when an error of list membership occurred, the within-list position judgments tended to remain accurate. This last finding contradicts an explicit prediction of strength theory; an item assigned to the wrong list should also be assigned to a within-list position that is temporally closest to its actual within-list position.

To explain these results, Hintzman postulated the existence of at least two types of contextual tags, each one referring to a particular subset of contextual elements.
The first type (Type A) referred to contextual elements that might, for example, define subjects' feelings of boredom during the encoding process. Hintzman further assumed that Type A elements change in a regular fashion according to a negatively function of time in an experiment. Therefore, Type A elements are assumed to change more quickly in the beginning of an experiment, and, thus, may explain the primacy effect. Moreover, if Type A tags change more quickly in the beginning of a presentation series than at the middle or end, then the positions of items in the beginning would be more discriminable than those from other portions.

Type A elements were also used to explain errors in list discrimination. These elements are presumably insensitive to list boundaries i.e., the rate at which they change does not correspond to the rate at which the stimulus series changes from one list to another. However, Type A elements retrieved with items from two consecutive lists might be quite similar to each other. It would, therefore, be reasonable to expect that list errors fall in lists that are in closest temporal proximity to the target list.

A second type of contextual marker (Type B) hypothesized by Hintzman reflects a subset of elements that describe the "cognitive environment" in which lists were processed. These elements may be associated with the degree of cognitive load or strain felt by subjects in the processing of each list. Type B elements were assumed to change in a regular
fashion during the presentation of a single list, but the original elements are reinstated whenever a new list is begun. Consequently, items from different lists, but with the same within-list position, share similar Type B elements (e.g., all beginning items might be associated with a minimum amount of strain, and end items a greater amount). Type B elements are used to explain the tendency for within-list judgments to remain accurate when errors in list discrimination are made.

On the basis of Hintzman's interpretations, certain tentative hypotheses can be formulated regarding the nature of serial list and position tags: (a) List and position tags are derived from contextual elements that describe subjects' internal, as opposed to external environment; (b) These elements must bear some logical correspondance with the passage of time (e.g., the greater the feeling of boredom, the longer the time spent in the experiment); (c) The discriminability of list tags are dependent upon the rate at which Type A elements change through time; the faster the rate of change, the more discriminable the tags. The rate of Type A change is, in turn, a function of time in a particular experimental setting; (d) The discriminability of position tags is dependent upon both the regularity with which Type B elements change within all lists (so that position tags are equivalent in all lists), and the degree to which Type B elements vary within each list (so that beginning tags are
different from middle and end tags).

The Present Experiment

Given Hintzman's interpretations of contextual tags, the present experiment was designed to study how certain variables of stimulus presentation would affect the discriminability of list and position tags in judgments of serial position. With regard to list tags, Hintzman hypothesized that the rate at which Type A elements change in an experiment is a negatively accelerated function of time in an experiment. This hypothesis, however, was offered as a post-hoc explanation for data collected in a study where certain "standard" conditions of stimulus presentation prevailed. Moreover, Hintzman's subjects viewed stimuli presented at a constant rate of duration (5 sec. per item), and with the same type of filler task (recognition memory test) intervening between lists. It could be argued that such conditions fostered a feeling of monotony or boredom (e.g., Type A elements) that might asymptote relatively early in the experiment. If presentation conditions were changed so that the experiment seemed less monotonous, it would be reasonable to expect list discrimination to improve. This would be a necessary correlate because the slope of the function describing Type A change would asymptote earlier in the experiment than under less monotonous conditions.

In the present study, three variables of stimulus pre-
sentation, presumed to influence contextual elements, were independently manipulated in a $2 \times 2 \times 2$ factorial design. The first variable was the type of filler task intervening between lists. In the Same task condition, subjects worked on an identical type of task between the presentation of each list. In the Different task condition, a different kind of task separated four presentation lists. It was felt that the Different condition would influence position knowledge in two ways. First, it might lessen the monotony of the experiment by the introduction of challenging, new tasks throughout the presentation series. Second, since each task represented a different cognitive activity, they might also provide subjects with a set of labels to differentially mark time during the presentation phase of the experiment. It was therefore hypothesized that list judgments in the Different condition would be better than those in the Same task condition.

The second variable manipulated in this study was total study time. In the Long time condition, subjects viewed each stimulus for an average of 8 seconds, in the Short condition, 5 seconds. The time interval between lists was held constant for both levels of this variable. Thus, the entire presentation phase for Long groups was 1.3 times longer than that for Short groups. It was hypothesized that the Short groups would do better in list discrimination than Long groups. In the Short condition, a greater proportion of stimuli would be
presented before any negatively accelerating function could asymptote. Thus, more of the items in the Short condition would be presented under a relatively "fast" rate of change than items in the Long condition. It can also be seen that the superiority of the Short study time should be most apparent in the beginning of the series (i.e., items from the latter part of the series would correspond with any asymptote of subjective experience).

The third variable manipulated was presentation pattern. As mentioned earlier, most studies of serial position have employed a constant rate of stimulus presentation. In the present experiment, the Constant condition of presentation pattern repeated this practice. In the Varied condition, on the other hand, subjects in the Short condition viewed stimuli for either 3, 5, or 7 seconds, so that an average of 5 seconds per item in each list was achieved. Subjects in the Long condition were shown stimuli for either 6, 8, or 10 seconds, resulting in an average of 8 seconds per item per list. There is considerable psychophysiological evidence to suggest that subjects' attentional state would be heightened under a Varied condition. Sokolov (1963), for example, has reported that when a long series of similar stimuli are shown at a constant rate, the physiological correlates of subjects' orienting response (i.e., attention) quickly habituate. However, when the duration of a stimulus changes from that of a preceding item, as in a varied condition, the orienting re-
sponse quickly recovers. As applied to tests of serial position, Sokolov's data implies that the Type A function, hypothesized by Hintzman, Block, and Summers (1973), would asymptote at a later point in the Varied condition than in the Constant condition. Therefore, it was hypothesized that a varied presentation pattern would result in better list discrimination than a constant one.

Of the three variables just described, only one was expected to have any influence on within-list position tags. Hintzman, Block, and Summers suggested that Type B elements reflect the degree of cognitive load or strain experienced during each lists' processing. There are probably a number of variables affecting the cognitive load associated with list processing. For example, the number of items in each list, the particular task demands, and the time it takes to process each list are only a few such variables. In the present experiment, it was reasoned that with the number of list items and task demands held constant, subjects in the Long condition would experience a greater degree of cognitive strain, particularly at the end of each list, than subjects in the Short condition. Moreover, in the Long condition, subjects would have to process each list over a much longer time interval than subjects in the Short condition. Thus, subjects in the Long condition would probably experience greater fatigue or "strain" at the end of each list than subjects in the Short condition. Consequently, there would be
a greater disparity between the Type B elements associated with beginning, middle, and end items in the Long condition. It was therefore hypothesized that the within-list position judgments in the Long condition would be more accurate than in the Short condition.
METHOD

Subjects

A total of 144 subjects, recruited from the undergraduate subject pool at Loyola University, were observed. Eighteen subjects were randomly assigned to each of the 8 experimental conditions.

Stimuli

Stimuli were 72 black-and-white photographs of objects likely to be seen in an urban environment (e.g., telephone booth, barber pole, etc.). All stimuli were screened beforehand to insure that none represented a landmark scene familiar to subjects.

In many studies of serial position knowledge, verbal stimuli have been used. There were two important reasons that prompted the use of pictures in the present study. First, it was essential that the stimuli have a high probability of being recognized during the test phase. According to the context model of memory, an item is not recognized unless contextual associates are retrieved. Since the present study was designed to investigate the effects of some variables on contextual tags, it was necessary that the position judgments reflect decisions for items recognized as having appeared in the experiment i.e., items whose contextual tags were retrieved. It is a well-known fact that pictures have a higher probability of recognition than words (Shepard, 1967; Standing,
Conezio, and Haber, 1970).

The second reason for using pictures stems from the fact that the present study was designed to observe groups with different study times. It was therefore necessary to insure that recognition memory did not vary with stimulus duration. Pavio (1971) reported that recognition memory for pictures does not vary with durations greater than 3 seconds. To verify this prediction, a pilot study was conducted in which 68 pictures, to be used in the present experiment, were viewed for two different durations. Half of the subjects (N = 20) saw each picture for 4 seconds, and the other half saw each picture for 8 seconds. The results confirmed Pavio's prediction. The hit rate for old items in a two-alternative forced choice recognition test was 93%, with no significant differences between long and short durations, F(1,18) = .334.

Presentation Lists

The 72 pictures were shown in four lists of 18 items each. Each of the three stimulus durations to be observed in the Varied condition were randomly assigned to positions within each list so that in each block of 6 items (corresponding to the beginning, middle, or end of each list) a different duration occupied each odd-numbered position. This restriction was imposed so that within each list, all stimulus durations were equally represented within all within-list positions (i.e., beginning, middle, and end), result-
ing in a constant study time for all four lists.

To control for any confounding effects due to the order of stimulus presentation, three different stimulus series were constructed according to the following criteria. First, each stimulus occupied either an odd-numbered or even-numbered position across all three stimulus series. This restriction followed from the fact that only odd-numbered stimuli served as test items, and it was essential that all subjects were tested on the same items. Second, the duration assigned to each stimulus in the Varied condition remained constant across each of the three series. Thus, the order in which the presentation rate varied in all three series remained the same (i.e., each of the durations in the Varied condition was assigned a series position, and in each of the three stimulus series, items were assigned positions that had the same duration). Third, each stimulus appeared in a different list in each series. Each stimulus series was equally represented in all experimental conditions.

**Test List**

Every stimulus occupying an odd-numbered position in the presentation series was used in the test series. Therefore, each list and within-list position in the presentation series was equally represented in the test list. To control for confounding effects due to test order, three different series were constructed so that: (a) Each block of 12 items
consisted of a beginning, middle, and end item from each of the four presentation lists; and (b) Within each block of 12 items, every duration used in the varied condition was equally represented.

Materials

Three types of filler tasks were chosen to fulfill three criteria: (a) Completion of each task did not require detailed instructions. Thus, subjects could spend a maximum amount of time working on the task, rather than focusing on task instructions; (b) Each type of task to be used in the Different condition involved a distinctly different cognitive activity. The first type was spatial (maze puzzle), the second type verbal (a vocabulary test used for graduate admissions), and the third type was mathematical (algebra problems); (c) Each type of task was challenging enough to prevent completion within the time interval allotted.

To record serial position judgments, subjects were asked to make a slash mark on a horizontal line (one for each test item) divided into four equal segments. Each consecutive segment corresponded to one of the four presentation lists. The segments corresponding to each list were further subdivided into three equal parts representing the beginning, middle, and end of each list. Thus, the accuracy of list and position judgments could be ascertained by noting whether the slash marks were placed in the appropriate
list and within-list segment.

Besides serial position judgments, the test sheets also contained a 5-point scale representing the confidence of subjects' judgments. If the subject felt he was guessing, he was to circle Number 1. All other numbers indicated ascending degrees of confidence, with Number 5 reflecting decisions for which subjects were "Very Sure."

Procedure

Subjects were seen in pairs and were given the following instructions prior to viewing the presentation series:

This is an experiment dealing with the relationship between human memory and certain mental aptitudes. You will be shown a fairly long series of pictures. Please pay close attention, because your memory for these pictures will be tested later in the experiment.

At various times throughout the picture series, the pictures will stop. When this happens, you are to begin work on a paper-and-pencil task that I will give to you. The instructions for each task are simple, and must be read quickly and silently. I will not entertain any questions either before or during each task. You will have a limited amount of time on each test, so work as quickly and efficiently as you can. When I tell you to stop, cease what you're doing immediately, and put the test face down in front of you. The pictures will then resume, and you will not return to the task once I've told you to stop. When the last picture in the series is shown, I'll give you further instructions.

Several aspects of these instructions should be noted. First, subjects did not know the nature of the serial position test during the presentation phase of the experiment. This was done to prevent subjects from using encoding strategies that may not generalize to real world settings. Sec-
ond, subjects were led to believe that the tasks represented measures of their mental aptitudes. By doing this, it was hoped that their involvement and attention with each task would be maximum.

Pictures in the presentation series were shown on a Kodak Carousel projector, and the presentation rates were automatically timed. The filler tasks were administered after the first, second, and third lists. Subjects were given 2 min. to work on each test. In the Same task condition, three different samples of one type of task were given, and each type of task (i.e., spatial, verbal, and mathematical) was used equally often in that condition. In the Different task condition, a sample of each of the three types of tasks were given. Three different orders of task administration were used equally often: spatial-verbal-math; math-spatial-verbal; verbal-math-spatial.

After the last list was presented, there was a 2 min. period prior to test in which answer sheets were distributed and test instructions given. During the test phase, each of the 36 test items was presented for 10 sec.
RESULTS

Serial position judgments were scored according to three criteria of accuracy, each criterion being treated as a separate dependent variable in the analyses. These criteria were: (a) List accuracy—was the list in which an item appeared remembered?; (b) Position accuracy—was the within-list position remembered (i.e., beginning, middle, or end), even though the item might have been placed in the wrong list?; (c) List-Plus accuracy—was both the list and within-list position of a test item recalled?

These criteria were used for two basic reasons. First, it was assumed that each criterion represented a different aspect of contextual retrieval. Successful list judgments presumably require the retrieval of list tags, position judgments require within-list position tags, and List-Plus accuracy requires both types of tags. Second, the criteria were viewed as representing varying degrees of task difficulty (with List-Plus the most difficult), and were therefore included to assess the degree to which the experimental variables might affect judgments differentially.

An examination of subjects' raw scores indicated that position knowledge was better than chance for each of the three criteria. For List accuracy, the mean number correct, out of a possible 36, was 16.5 (chance = 9.0). The corresponding means for the Position and List-Plus criteria were 14.0 (chance = 12.0), and 7.0 (chance = 3.0) respectively
Subjects' raw scores were transformed to response probabilities. Moreover, subjects' list and position judgments were not evenly distributed among all possible response categories. With regard to list assignments, for example, subjects assigned items to Lists 2 and 3 far more frequently than to Lists 1 or 4. Also, the within-list assignments were biased in favor of the middle positions (see Table 2).

The uneven distribution among response categories was interpreted to mean that subjects exhibited a guessing bias for this particular task. Thus, to appropriately test the hypotheses of the present experiment, it was first necessary to correct each subject's score for his particular guessing bias (e.g., the raw number correct of list judgments for List 3 items might be greater than that for List 1; but such a difference would be due to the fact that when guessing, subjects tend to assign items to List 3 rather than to List 1). A correction for guessing was used in which the raw scores for each of the scoring criteria were transformed by dividing the raw number correct for a given list and/or position, by the frequency with which a subject used the list and/or position as a response category. This resulted in post-hoc probabilities of a correct response (cf. Hintzman et al., 1973).

Following the above transformations, the response probabilities for each of the three scoring criteria were separ-
<table>
<thead>
<tr>
<th>Group</th>
<th>List M</th>
<th>List SD</th>
<th>Position M</th>
<th>Position SD</th>
<th>List-Plus M</th>
<th>List-Plus SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Long</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Same-Con.</td>
<td>16.6 (3.6)</td>
<td></td>
<td>13.8 (3.0)</td>
<td>7.9 (2.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Same-Var.</td>
<td>16.3 (3.4)</td>
<td></td>
<td>14.1 (3.5)</td>
<td>7.2 (2.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diff-Con.</td>
<td>16.2 (3.5)</td>
<td></td>
<td>14.6 (2.2)</td>
<td>6.9 (2.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diff-Var.</td>
<td>15.8 (3.7)</td>
<td></td>
<td>14.0 (3.0)</td>
<td>7.5 (3.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Same-Con.</td>
<td>16.6 (3.2)</td>
<td></td>
<td>14.6 (3.0)</td>
<td>7.3 (2.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Same-Var.</td>
<td>17.8 (4.2)</td>
<td></td>
<td>13.3 (3.6)</td>
<td>7.1 (2.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diff-Con.</td>
<td>15.3 (3.0)</td>
<td></td>
<td>13.9 (3.5)</td>
<td>6.1 (1.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diff-Var.</td>
<td>17.0 (2.2)</td>
<td></td>
<td>13.2 (2.6)</td>
<td>6.9 (2.2)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
TABLE 2

RESPONSE FREQUENCY OF CATEGORIES\(^a\)

<table>
<thead>
<tr>
<th>Position</th>
<th>B</th>
<th>M</th>
<th>E</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>List 1</td>
<td>.047</td>
<td>.050</td>
<td>.058</td>
<td>.153</td>
</tr>
<tr>
<td>List 2</td>
<td>.066</td>
<td>.128</td>
<td>.093</td>
<td>.288</td>
</tr>
<tr>
<td>List 3</td>
<td>.108</td>
<td>.141</td>
<td>.107</td>
<td>.354</td>
</tr>
<tr>
<td>List 4</td>
<td>.081</td>
<td>.094</td>
<td>.025</td>
<td>.201</td>
</tr>
<tr>
<td>Mean</td>
<td>.304</td>
<td>.412</td>
<td>.284</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) Percentage scores
ately analyzed within an appropriate ANOVA design. It should be noted, however, that before these analyses were conducted, the control variables of presentation series and task order were tested for any main effects or interactions. No significant differences between series, $F(2, 63) = 1.23; p. .30$, or task order, $F(2, 63) = .23$, were found.

**List Accuracy**

To test for list accuracy, the probabilities of a correct list response for items from each list were computed, and analyzed within a $2 \times 2 \times 2 \times 18 \times 4$ ANOVA design (Task x Presentation Pattern x Time x Subjects x List); with subjects nested within the Task x Pattern x Time interaction, and repeated across lists. A main effect for List was found, $F(3, 408) = 200; p. .001$; with items from Lists 1 and 4 resulting in better discrimination than those from 2 or 3 (see Figure 1).

There were no main effects for any of the between-group variables. However, by comparing the upper panels with the lower panels in Figure 1, an interaction between Time and Pattern is indicated. The Varied condition enhanced list discrimination for the Short time relative to the Long time. This interaction reached a level of statistical significance, $F(1, 136) = 4.34; p. .04$. A simple effects analysis for the variable of Pattern showed that the differences between Constant and Varied groups in the Short condition approached sig-
Figure 1

Plot showing the probability of correct response against list for both short and long lists with 'same' and 'different' conditions. The graph includes lines for constant and varied conditions.
nificance, $F(1, 136) = 3.48; p. .06$.

**Position Accuracy**

To test for position accuracy, the correct response probabilities for each within-list position (i.e., beginning, middle, and end positions summed across all Lists) were computed and analyzed within a $2 \times 2 \times 2 \times 18 \times 3$ ANOVA design (Task x Pattern x Time x Subjects x Position); with subjects nested in the Task x Pattern x Time interaction, and repeated across Positions. A main effect for Position was found, $F(2, 272) = 30.9; p. .001$; beginning items having better accuracy than middle or end items.

Figure 2 describes the probabilities of correct position judgments for all the experimental groups. By comparing the upper panels with the lower panels, an interaction between Time and Position is suggested. Under the Long condition, end items were better recalled than middle items; whereas, under the Short time, end items were poorer than middle items, $F(2, 272) = 3.24; p. .04$. A simple effects analysis for the variable of Time showed that for end items, Long groups were significantly better than Short groups, $F(1, 408) = 6.66; p. .01$.

The position data were examined in more detail by separating the mean correct response probabilities for each position into two components: (a) the probability of a correct position response in the right list; and (b) the probability...
of a correct position response in the wrong list (see Table 3). From Table 3, it is clear that subjects tended to remember an item’s within-list position, even when errors in list discrimination occurred. Note that for middle and end items, the response probabilities in the wrong list were generally equal to or higher than those in the right list. Such a pattern could not occur unless subjects were able to retrieve position tags somewhat independently of list tags. Clearly, Hintzman et al.’s (1973) finding (i.e., the tendency for correct position responses in wrong list) for verbal stimuli, was replicated for picture stimuli in the present experiment.

List-Plus Accuracy

To test for List-Plus judgments, the response probabilities for items from each of the three positions within each list were computed and analyzed within a $2 \times 2 \times 2 \times 18 \times 4 \times 3$ ANOVA design (Task x Pattern x Time x Subjects x List x Position); with subjects nested in the Task x Pattern x Time interaction, and repeated across the List x Position interaction. Main effects for List, $F(3, 408) = 91.9$; p. .001; and Position, $F(2, 272) = 76.9$; p. .001, were found. Items from Lists 1 and 4 were assigned more accurately than those from Lists 2 or 3, and the correct response probabilities for beginning items were better than those for middle or end items.

From Figure 3, which depicts the response probabilities
TABLE 3

POSITION DATA INTO SEPARATE COMPONENTS

Probabilities of Correct Position Response

<table>
<thead>
<tr>
<th></th>
<th>Right List</th>
<th></th>
<th>Wrong List</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>M</td>
<td>E</td>
<td>B</td>
</tr>
<tr>
<td>Groups</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Long</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Same-Con.</td>
<td>.28</td>
<td>.16</td>
<td>.15</td>
<td>.16</td>
</tr>
<tr>
<td>Same-Var.</td>
<td>.27</td>
<td>.16</td>
<td>.13</td>
<td>.17</td>
</tr>
<tr>
<td>Diff-Con.</td>
<td>.25</td>
<td>.13</td>
<td>.21</td>
<td>.21</td>
</tr>
<tr>
<td>Diff-Var.</td>
<td>.27</td>
<td>.16</td>
<td>.23</td>
<td>.16</td>
</tr>
<tr>
<td>Mean Long</td>
<td>.27</td>
<td>.15</td>
<td>.18</td>
<td>.18</td>
</tr>
<tr>
<td>Short</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Same-Con.</td>
<td>.30</td>
<td>.16</td>
<td>.15</td>
<td>.20</td>
</tr>
<tr>
<td>Same-Var.</td>
<td>.29</td>
<td>.18</td>
<td>.14</td>
<td>.15</td>
</tr>
<tr>
<td>Diff-Con.</td>
<td>.27</td>
<td>.14</td>
<td>.11</td>
<td>.20</td>
</tr>
<tr>
<td>Diff-Var.</td>
<td>.25</td>
<td>.18</td>
<td>.17</td>
<td>.17</td>
</tr>
<tr>
<td>Mean Short</td>
<td>.28</td>
<td>.16</td>
<td>.14</td>
<td>.18</td>
</tr>
<tr>
<td>Total Mean</td>
<td>.275</td>
<td>.155</td>
<td>.16</td>
<td>.18</td>
</tr>
</tbody>
</table>
FIGURE 3A

SHORT-SAME

SHORT-DIFFERENT

--- CONSTANT --- VARIED
FIGURE 3B

LONG-SAME

LONG-DIFFERENT

B M E B M E B M E B M E

1 2 3 4

--- CONSTANT ---- VARIOUS
for all of the experimental groups, it can be seen that in 7 of the 8 groups, beginning items from List 1 (1B), and end items from List 4 (4E) have the highest correct response probabilities. Thus, for the most part, the primacy and recency effects reported for verbal stimuli (e.g., Hintzman et al., 1973) was further replicated in the present experiment for pictures. The only exception to that finding is seen in the condition of Short time, with Different tasks and a Constant rate (second panel of Figure 3).

Confidence

In addition to observing accuracy, the confidence of each subject's decision for each list and within-list position was analyzed within a design described for List-Plus accuracy. For the most part, subjects' degree of confidence corresponded to their degree of response accuracy; the higher the correct response probability, the greater the confidence. Table 4 shows the mean confidence judgments for each position within each list. Main effects for List, $F(3, 408) = 69.7; p. .01$; and Position, $F(2, 272) = 13.1; p. .01$, were found; as well as a List x Position interaction, $F(6, 816) = 57.3; p. .001$.

Effects of Duration

An analysis was also performed to test for any main effects or interactions for the variable of stimulus dur-
<table>
<thead>
<tr>
<th>List</th>
<th>B</th>
<th>M</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>List 1</td>
<td>4.06</td>
<td>3.10</td>
<td>2.88</td>
</tr>
<tr>
<td>List 2</td>
<td>2.69</td>
<td>2.75</td>
<td>2.79</td>
</tr>
<tr>
<td>List 3</td>
<td>2.67</td>
<td>2.80</td>
<td>2.86</td>
</tr>
<tr>
<td>List 4</td>
<td>2.88</td>
<td>2.94</td>
<td>3.29</td>
</tr>
</tbody>
</table>
ation. As mentioned earlier, in Varied groups stimuli were presented for one of three durations, with 4 sec. between the shortest and longest durations. To test for the effects of stimulus duration, the raw number correct for each duration was computed according to each of the three scoring criteria. These data were then analyzed within a 2 x 2 x 18 x 3 ANOVA design (Task x Time x Subjects x Duration), with subjects nested in the Task x Time interaction, and repeated across Duration. On all three criteria, a main effect for duration was found: List accuracy, \( F(2, 136) = 3.45; p = .03 \); Position accuracy, \( F(2, 136) = 10.7; p = .005 \); List-Plus, \( F(2, 136) = 19.0; p = .001 \). From Table 5, it can be seen that the long durations resulted in more accurate judgments than medium or short durations.

Of particular interest in Table 5 is the absence of any main effects for total study time. Moreover, there were no consistent differences between Long and Short scores for each of the durations (recall that durations in Long were 6, 8, and 10 sec., and 3, 5, and 8 sec. in the Short condition). For example, on list discrimination, the scores of the Short groups tended to be better than those in the Long. This corresponds to the previously reported finding that the Varied rate enhanced list judgments in the Short, but not in the Long condition. For position accuracy, on the other hand, Long groups tended to do better than Short groups, a finding congruent with the fact that the Long condition facilitated
TABLE 5

ACCURACY FOR ITEMS OF EACH DURATION
(Mean Number Correct)

<table>
<thead>
<tr>
<th></th>
<th>Duration</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>S</td>
<td>M</td>
</tr>
<tr>
<td>List</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Long</td>
<td>4.91</td>
<td>5.33</td>
</tr>
<tr>
<td>Short</td>
<td>5.55</td>
<td>6.08</td>
</tr>
<tr>
<td>Mean</td>
<td>5.23</td>
<td>5.70</td>
</tr>
<tr>
<td>Position</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Long</td>
<td>4.16</td>
<td>4.33</td>
</tr>
<tr>
<td>Short</td>
<td>4.41</td>
<td>4.02</td>
</tr>
<tr>
<td>Mean</td>
<td>4.29</td>
<td>4.18</td>
</tr>
<tr>
<td>List-Plus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Long</td>
<td>1.88</td>
<td>2.16</td>
</tr>
<tr>
<td>Short</td>
<td>2.05</td>
<td>2.16</td>
</tr>
<tr>
<td>Mean</td>
<td>1.97</td>
<td>2.16</td>
</tr>
</tbody>
</table>
performance for end items. Thus, it appears that time per se was not responsible for the duration effects. Rather, variation of stimulus duration seemed to affect subjects in a relative fashion, within the context of a particular total study time.

Mean Placement of Items

Finally, an analysis was conducted to determine the effects of duration upon the mean placement of test items. Recall that each test item could be assigned to any one of 12 line segments, representing each within-list position of every presentation list. Depending on where a subject assigned a test item, his slash mark on the answer sheet was translated into a number corresponding to the particular line segment (i.e., 1 = beginning of List 1, 2 = middle of List 1, 4 = beginning of List 2, etc.). These numbers were then analyzed within a 2 x 2 x 18 x 3 ANOVA design (Task x Time x Subjects x Duration); with Subjects nested in the Task x Time interaction and repeated across Duration.

This analysis bears directly upon a prediction of strength theory; items of longer duration should result in stronger traces, and should, therefore, be assigned to more recent serial positions than items of shorter duration. Strength theory would also expect a main effect for total study time, with Long groups assigning items to more recent serial positions than Short groups. Contrary to these predictions,
there were no main effects for total study time; and in Varied groups, items presented for the shortest durations (i.e., either 3 or 6 sec., depending upon the total study time) were assigned to more recent, rather than less recent serial positions, $F(2, 136) = 4.85$; $p. .009$ (see Table 6).
TABLE 6
MEAN PLACEMENT FOR ITEMS OF EACH DURATION

<table>
<thead>
<tr>
<th>Groups</th>
<th>S</th>
<th>M</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Same-Long</td>
<td>7.09</td>
<td>7.00</td>
<td>6.62</td>
</tr>
<tr>
<td>Diff-Long</td>
<td>7.06</td>
<td>6.75</td>
<td>6.62</td>
</tr>
<tr>
<td>Mean-Long</td>
<td>7.07</td>
<td>6.87</td>
<td>6.62</td>
</tr>
<tr>
<td>Same-Short</td>
<td>6.93</td>
<td>6.90</td>
<td>6.68</td>
</tr>
<tr>
<td>Diff-Short</td>
<td>6.82</td>
<td>6.84</td>
<td>6.76</td>
</tr>
<tr>
<td>Mean-Short</td>
<td>6.88</td>
<td>6.87</td>
<td>6.72</td>
</tr>
<tr>
<td>Total Mean</td>
<td>6.98</td>
<td>6.87</td>
<td>6.66</td>
</tr>
</tbody>
</table>
DISCUSSION

Of the four hypotheses tested in this experiment, only two received some empirical support. The first one predicted that Varied groups would be better list discriminators than Constant groups. It was found, in line with this prediction, that a varied rate enhanced performance, but only under the condition of Short study time. Failure to find the predicted effect under the Long time can be explained by recalling Weber's law: as the absolute value of a stimulus increases, sensitivity to change in the stimulus decreases. The variation in stimulus duration in the Long condition (6, 8, or 10 sec.) was probably not as noticeable as the variation in the Short condition (3, 5, or 7 sec.). In support of this position, it should be mentioned that many of the subjects in the Short-Varied groups commented that they were aware of the variation in duration; whereas, few, if any, of the the subjects in the Long-Varied groups did the same.

In the present experiment, it was hypothesized that a varied rate would enhance list discrimination through its effects upon the subjects' attentional state; which, in turn, would influence the slope of Hintzman's hypothetical Type A function. However, in view of the additional finding that list discrimination improved with stimulus duration, it might be argued that the varied rate enhanced performance only because, in that condition, some stimuli were presented for a longer time than in the Constant condition. This counter-argument can be rejected for at least two reasons. First,
in the Varied condition, there were as many stimuli with durations shorter than those in the Constant, as there were stimuli with longer durations. Thus, the beneficial effects of the longer durations in the Varied condition would have been cancelled by the equal number of shorter durations. Second, if the varied rate enhanced performance only because of longer durations, the effect should also have been observed in the Long condition. Clearly, the data support the position that a varied rate helps list discrimination because of its effects on subjects' attentional state.

The other hypothesis to receive empirical support was one that predicted that subjects in the Long condition would show more accurate within-list position judgments than subjects in the Short condition. It was reasoned that a long study time would increase subjects' experience of cognitive strain during the processing of each list, thereby improving the discriminability of position tags (i.e., by increasing the disparity between the degree of load associated with beginning, middle, and end items). Consistent with this prediction, it was found that Long groups were more accurate in their judgments for end items than Short groups. Of particular interest, however, was the failure to find similar group differences for beginning and middle items. Moreover, if the Long condition increased the experience of cognitive strain, one would have also expected the discrimination between beginning and middle positions to improve as well.
Why did the Long time affect end items only? A possible explanation can be found by recalling subjects' response biases shown in Table 2. For reasons not clearly understood, subjects assigned test items to response categories representing end positions much less frequently than they did to middle or beginning positions. This could be interpreted to mean that criteria, implicitly used by subjects to define and describe end positions, were much more stringent and precise than those for beginning or middle positions. It is also helpful to recall that "end" items were arbitrarily defined as those stimuli occupying the 13th, 15th, and 17th positions within each list. Thus, it was very difficult to distinguish middle items (7th, 9th, and 11th positions) from some of the end items. In the Short condition, the degree of cognitive strain associated with end items was probably not great enough to warrant their assignment to an end response category. In the Long condition, subjects probably experienced an added degree of strain while encoding end items, and this added amount would help them distinguish end positions from middle positions.

Besides affecting within-list position judgments, the variable of Time was also hypothesized to affect list discrimination. According to Hintzman, Block, and Summers (1973), the amount of time per se subjects spend in an experimental setting is a crucial factor in the accuracy of list judgments. Presumably, the longer the time spent in a setting, the more
likely it is that subjects will encode stimuli within a psychological context (i.e., feelings of boredom) whose elements have reached an asymptote of change. Thus, it was predicted that list judgments would be better under the condition of a relatively short presentation time. The present experiment, however, failed to confirm this hypothesis; the variable of study time had no effect on list judgments.

It could be argued that in the present experiment, the differences between the Long and Short presentation times were not large enough to show any noticeable differences in list discrimination. Recall that the presentation phase (including task intervals) in the Long condition lasted for 15.6 min., or approximately 1.3 times longer than the corresponding time in the Short condition (12 min.). Obviously, further research could test this argument by making the time differences between Long and Short groups greater. However, the present author has concluded that the variable of total time per se will have little effect, even if the differences between Long and Short groups were more extreme. Moreover, from Hintzman's assumptions, it follows that the superiority of a shorter time should be most apparent in the beginning of a series. Thus, in the present experiment, Short groups should have performed better than Long groups in at least the first two lists. The data, however, do not agree with this expectation. In fact, the best performance for List 1 items was given by a group under the Long condition (see Figure 1;
Long-Same-Varied with .81 mean correct response probability).

Rather than time per se in an experimental setting, it might be better to consider subjects' experience of time spent in a setting, as a crucial factor in list discrimination. Some psychologists have recently advanced the hypothesis that a person's sense of duration is dependent, at least in part, upon the number of stimuli he encodes in a finite interval of time. Ornstein (1974) and Block (1974) have both demonstrated that a time interval filled with the presentation of 80 stimuli appears to be longer than one filled with 40 stimuli. In the present experiment, the number of items encoded by the Long and Short groups were identical, and it is therefore conceivable that the experience of temporal duration in both groups was comparable. Thus, the failure to find differences between these groups in the present study might be explained by considering Ornstein's data. If the experience of duration was equivalent in both groups, the hypothetical Type A function probably reached asymptote at similar points along the presentation series in both the Long and Short conditions.

Failure to confirm the fourth experimental hypothesis, that predicted better list discrimination in the Different task condition, necessitates yet another revision of Hintzman's assumptions regarding the Type A function. As explained earlier, Hintzman assumed that the function of Type A contextual change extended throughout the presentation phase
of an experiment. This implied that subjects' activities between, as well as within lists, would affect the slope of the function, and, thus, list discrimination. Contrary to that position, however, the present experiment found no evidence that subjects' interlist activities affected list judgments. The subjective experience (e.g., feelings of boredom) associated with stimulus lists was the same, whether a novel or repetitive task intervened between them.

This latter finding is rather hard to understand intuitively, especially in view of the fact that subjects were under the erroneous impression that the filler tasks were to be used as measures of their mental aptitudes. Most all of the subjects in the Different task condition appeared to be highly motivated during each type of task, and many of them expressed concern over their failure to complete the tests during the 2 min. intervals. In the Same task condition, however, many subjects reported that after the task type was repeated for the second time, they began to suspect their true function in the experiment, and were, therefore, not as concerned about their performance. For this reason alone, one would have expected the different tasks to have had some effect on list judgments. Apparently, no matter how motivated or involved subjects might have been with filler tasks, each succeeding presentation list was experienced in a similar fashion.

However, it is quite possible that the manipulation of
Different tasks in the present experiment was not enough to alter subjects' psychological context. Moreover, it could be that the subjects treated the filler tasks the same, regardless of their nature. Perhaps a more powerful manipulation would entail the comparison of Task/No-Task groups. Underwood (1969), for example, cited several studies to show that subjects allowed to engage in various physical activities between lists (e.g., walking out of the experimental room for a drink of water) showed reduced inter-list interference on a paired-associate learning task. The reduced interference between lists was interpreted to mean that those subjects had established better temporal discrimination between the lists. As applied to the context hypothesis, this finding would lead to the prediction that various physical activities between lists would result in better list discrimination than filler tasks between lists.

There was another finding in the present study that was totally unexpected. In the Different-Short-Constant condition, subjects failed to show the usual strong recency effect for 4E items. This was particularly surprising, since previous research has shown the recency effect to be quite pervasive (e.g., Guenther and Marigold, 1975; Hintzman and Block, 1971; Underwood, 1969). Until now, the only way known to attenuate this effect was by prolonging the time interval between the presentation of the last series items, and the time of test (e.g., Guenther and Marigold, 1975; Hintzman et al., 1973).
Hintzman et al. (1973) hypothesized that at short retention intervals, subjects infer additional recency information for items at the end of a series from the degree to which contextual elements associated with these items match contextual elements prevailing at the time of test. At longer retention intervals, however, the degree to which elements retrieved with test items match prevailing contextual elements would be reduced. Thus, the attenuation of the recency effect at longer intervals was explained.

In the present experiment, however, the retention interval between the presentation and test phases of the session were identical for all experimental groups. Thus, Hintzman's hypothesis would be hard-pressed to explain why the Short-Different-Constant group failed to show the strong recency effect. No immediate explanation for this finding can be offered by the present author. However, the results clearly indicate that Hintzman's original interpretation for recency effects must be re-examined.

Present Findings in Perspective

The results of the present experiment indicate that two of Hintzman's assumptions regarding the nature of contextual change (i.e., Type A elements) need to be revised or modified. First, time per se in an experimental setting does not appear to be the determining factor in the slope of a Type A function. Rather, time as experienced in a situation
might be a better way of conceptualizing the abscissa of a hypothetical Type A function. Second, Type A elements associated with list items appear to be unaffected by the nature of inter-list tasks. It is quite possible that other types of inter-list activities (i.e., activities that do not entail paper-and-pencil tests) could alter the slope of a Type A function, but further research is needed to determine this.

Although the above results contradict some assumptions regarding Type A change, they should not be interpreted to mean that all of Hintzman's ideas about contextual change are wrong. Quite the contrary, the present experiment did observe enhanced list discrimination with a varied rate (i.e., marginally significant under a Short time), and this prediction followed directly from Hintzman's premise that subjects attentional state (or "feelings of boredom" in Hintzman's terms) is of fundamental importance in the formation of discriminable list tags. Furthermore, the fact that within-list position judgments appear to operate somewhat independently of list judgments agrees with Hintzman's notion that there are separate list (Type A elements) and position (Type B elements) tags. Finally, the finding that Long groups performed better for end items than Short groups is congruent with the idea that Type B elements might reflect the degree of cognitive load or strain associated with stimuli.

Of particular interest in the present study, was the
finding that in both Long and Short Varied groups, items of shortest duration were assigned to more recent serial positions than items of longer duration. As mentioned earlier, this result contradicts an explicit prediction of strength theory. The context hypothesis, however, is able to offer the following post-hoc explanation for these data. With relative increases in an item's duration, the greater the likelihood that Type A and B elements associated with the item change during its presentation. (This would, by the way, also explain why the accuracy of judgments improved with relative increases in duration.) When a subject retrieves an item's contextual elements at the time of test, he might implicitly realize that some items are associated with a greater degree of contextual change than others. That is, items of longer duration would be associated with more contextual change than items of shorter duration. Because of this implicit recognition, subjects might have the tendency to assign longer items to less recent response categories. Moreover, according to Hintzman's interpretations, items are correctly assigned to the beginning of a series (i.e., the primacy effect) because they were encoded in the midst of greater contextual change. It could be that items of longer duration are assigned to earlier serial positions for the same reason.

Conclusions

In light of the above discussion, the present author
concludes that Hintzman's interpretations of the context hypothesis offers the best possible explanation for serial position knowledge that is currently available. It should be emphasized, however, that the Type A and Type B elements hypothesized by Hintzman are not meant to serve as explanations for all types of recency phenomena. Rather, Type A and B elements appear to be the most likely mediators of contextual tags in tests of serial position.
REFERENCES


Shepard, R. Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, 1967, 6, 156-163.


Underwood, B.J. Attributes of memory. *Psychological Review*,


APPROVAL SHEET

The thesis submitted by Kathleen A. Carlson has been read and approved by the following committee:

Dr. Eugene B. Zechmeister, Director
Associate Professor, Psychology, Loyola

Dr. Frank L. Slaymaker
Assistant Professor, Psychology, Loyola

Dr. David J. De Palma
Assistant Professor, Psychology, Loyola

The final copies have been examined by the director of the thesis and the signature which appears below verifies the fact that any necessary changes have been incorporated and that the thesis is now given final approval by the Committee with reference to content and form.

The thesis is therefore accepted in partial fulfillment of the requirements for the degree of Master of Arts.

12-8-75
Date

Director's Signature