Imagery and Syntactic Reconstruction of Sentences

Patrick James Carroll

College of William & Mary - Arts & Sciences

Follow this and additional works at: https://scholarworks.wm.edu/etd

Part of the Cognitive Psychology Commons, and the Linguistics Commons

Recommended Citation


https://dx.doi.org/doi:10.21220/s2-n9n0-jz54

This Thesis is brought to you for free and open access by the Theses, Dissertations, & Master Projects at W&M ScholarWorks. It has been accepted for inclusion in Dissertations, Theses, and Masters Projects by an authorized administrator of W&M ScholarWorks. For more information, please contact scholarworks@wm.edu.
IMAGERY AND SYNTACTIC RECONSTRUCTION OF SENTENCES

A Thesis
Presented to
The Faculty of the Department of Psychology
The College of William and Mary in Virginia

In Partial Fulfillment
Of the Requirements for the Degree of
Master of Arts

by
Patrick J. Carroll
1977
APPROVAL SHEET

This thesis is submitted in partial fulfillment of
the requirements for the degree of
Master of Arts

Cynthia H. Null, Ph.D., Chairperson
Peter L. Derks, Ph.D.

Approved, August 1977

Cynthia H. Null, Ph.D., Chairperson
Peter L. Derks, Ph.D.

Deborah G. Hartley, Ph.D.

E. Rae Harcum, Chairperson
Department of Psychology
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acknowledgments</td>
<td>iv</td>
</tr>
<tr>
<td>List of Tables</td>
<td>v</td>
</tr>
<tr>
<td>Abstract</td>
<td>vi</td>
</tr>
<tr>
<td>Introduction</td>
<td>2</td>
</tr>
<tr>
<td>Pretesting</td>
<td>11</td>
</tr>
<tr>
<td>Method</td>
<td>22</td>
</tr>
<tr>
<td>Results and Discussion</td>
<td>31</td>
</tr>
<tr>
<td>Conclusions</td>
<td>46</td>
</tr>
<tr>
<td>Appendix</td>
<td>51</td>
</tr>
<tr>
<td>References</td>
<td>64</td>
</tr>
</tbody>
</table>
ACKNOWLEDGMENTS

The author wishes to express his appreciation to Cynthia Null, Peter Derks and Deborah Hartley for their patience, encouragement and constructive criticism during the birth, death, and resurrection of this study. The author would also like to thank his friends and especially his parents who helped revive sagging spirits through their support and encouragement. Finally, a very personal thanks to Nancy Buckley for caring and never giving up. This study was partially supported by a grant awarded in November, 1974, by the Faculty Research Committee of the College of William and Mary.
<table>
<thead>
<tr>
<th>No.</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Imagery ratings (7-point scale) of the paragraph contexts with key sentences present (Pretest I) and absent (Pretest II)</td>
<td>15</td>
</tr>
<tr>
<td>2.</td>
<td>Meaningfulness ratings (7-point scale) of the paragraph contexts with key sentences present (Pretest I) and absent (Pretest II)</td>
<td>17</td>
</tr>
<tr>
<td>3.</td>
<td>Pretest III: Mean ratings (7-point scale) of key sentence meaningfulness according to sentence type and contextual imagery</td>
<td>18</td>
</tr>
<tr>
<td>4.</td>
<td>Proportion of the ratings indicating that the logical recipient of the key sentence was the theme of the context</td>
<td>21</td>
</tr>
<tr>
<td>5.</td>
<td>Percent of the stimulus sentences recalled, combining the correct and transformed sentences</td>
<td>35</td>
</tr>
<tr>
<td>6.</td>
<td>Total number of key sentences recalled in original stimulus voice</td>
<td>37</td>
</tr>
<tr>
<td>7.</td>
<td>Total number of key sentences transformed in voice</td>
<td>38</td>
</tr>
<tr>
<td>8.</td>
<td>Transformational probabilities of key sentences comparing James, Thompson, and Baldwin (1973) and the present experiment</td>
<td>40</td>
</tr>
<tr>
<td>9.</td>
<td>Total number of nouns recalled</td>
<td>44</td>
</tr>
</tbody>
</table>
ABSTRACT

The fact that retrieval is an active process which can distort the actual contents of memory has been given insufficient attention by language theorists. The "reconstructive hypothesis," a recall theory, of James, Thompson, and Baldwin (1973) is here attached to two popular theories of memory storage, the imagery theory of Paivio (1971a) and the "constructive hypothesis" of Bransford, Barclay, and Franks (1972). The purpose is both to test these storage theories and to re-evaluate the "reconstructive hypothesis" in light of the possibility that it is interpreting as the effects of imagery levels what is in fact a difference in meaningfulness among nouns.

The experiment was carried out in a cued-recall situation using sentences with various combinations of noun imagery as stimuli. A preceding paragraph placed each sentence in a context, thus reducing the meaningfulness problem. The paragraphs also served as recall cues.

The results showed that the "reconstructive hypothesis" accurately predicted the probability of voice transformation among sentences with different noun-imagery combinations, the most important statistic, but that, contrary to earlier findings, memory was as good for low imagery as for high imagery items. Results were interpreted as supportive of the "constructive hypothesis" of storage. The "reconstructive hypothesis" of recall was seen as basically correct except for its dependency on the hypothetical notion of imagery. A new hypothesis is proposed to account for the selective effects of stimuli rated according to imaginal value.
IMAGERY AND SYNTACTIC RECONSTRUCTION
OF SENTENCES
INTRODUCTION

Many influential theories of language comprehension are inadequate because, while they take into account the dynamic nature of understanding, they depend upon a naive view of retrieval of information from memory. This is the position taken by James, Thompson, and Baldwin (1973) as the starting point for a study which tested several current storage theories by supplementing them with a more sophisticated retrieval hypothesis that they previously included. Their strategy is interesting especially because it provides a convenient way of setting up competing theories in a situation where they may predict different outcomes. The retrieval process acts as a sort of constant and mediator between the hidden contents of memory and the tangible products of recall. The step is possible for theories having no strong commitment to a particular retrieval process.

The potentially weak link in this strategy is the retrieval hypothesis. To the extent that it is inaccurate, the avowed purpose of distinguishing among storage theories will be foiled. This paper will re-examine a crucial feature of the James, et al. theory (hereafter referred to by their title: the "Reconstruction Hypothesis"), its dependence upon the memorial effects of visual imagery. Imagery, by their account, affects memory by variations in "vivacity" of the cognitive representation. Imagery is, unfortunately, easily confounded with meaningfulness or comprehensibility of individual words. By placing words in sentences and sentences in a larger context, some of
the differences usually found in meaningfulness between abstract and concrete words can be eliminated.

The plan of the Introduction is as follows. We will first review two storage theories which disagree about the contents of memory and, consequently, lead to different predictions about what will be recalled. Next, the Reconstruction Hypothesis will be explained in its original form, including a look at the causes of some response tendencies (voice and theme) and a brief review of the experiment which is the test of their predictions. Familiarity with the two storage theories will permit reconstruction to be seen as it was intended: as a consequence of the structure and the contents of memory rather than as an isolated event. The Reconstruction Hypothesis will then be evaluated in light of the suggestion that meaningfulness and imagery have frequently been confused in experiments on imagery. Finally, predictions following from the combination of the various storage and recall theories will be stated explicitly.

The storage theories. Paivio's "two-process model" of linguistic representation is doubly important to James and his colleagues. It is one of the memorial theories on which they focus and the functions of imagery which it describes are considered to be in recall quite independent of the overall model's validity. For now, we shall just look at its storage aspect (see Paivio, 1971a, for a general review of the theory and of its empirical support). Paivio considers "modes of representation" to be a basic cognitive dimension which, while perhaps not the sole residence of meaning, is certainly a major determinant of the form in which a word or larger language unit will be understood and later used. Concrete or high imagery items are coded
and stored in a non-verbal or imaginal form while abstract or low imagery items are processed in a verbal form related directly to speech. The boundaries between these cognitive modes are not absolute, particularly for concrete words; but, as Paivio (1971b) makes clear, the stimulus-based tendency to use one mode or the other is compelling and easily measurable in verbal behavior. (The terms high imagery and concrete and the terms low imagery and abstract will be used interchangeably throughout this paper.) On the level of individual words, imagery aids memory by acting as a vivid recall cue, as can be seen by its use in traditional mnemonic strategies (Paivio, 1969; Paivio, Yuille, and Smythe, 1966). A more complex relationship exists for words in sentences. High imagery value facilitates memory for meaning and low imagery value helps memory for syntax (Begg and Paivio, 1969; Begg, 1971). They suggest that this occurs because the words in a concrete sentence are represented as a sequential verbal string while the meaning is a non-verbal spatial image. Meaning is normally all that is important, hence, all else is discarded from memory. Even on the level of a paragraph, memory for individual words is aided by the mediation of a thematically integrated, high imagery context (Yuille and Paivio, 1969).

The other storage theory is that of Bransford, Barclay, and Franks (1972) who, following Fillenbaum (1966) and Sachs (1967), feel that language in its public grammatical form is quickly discarded by the language processor. Their theory, which they call the "constructive approach to sentence memory," argues that sentences merely act as cues to allow people "to recreate and modify their previous knowledge of the world [p. 207]." Unless there is a pressing reason to retain syntactic structures, these are quickly discarded. All that
remains may be what Fillenbaum (1966) called the "gist," the general idea which includes some logical relations. Recall will be consonant in meaning with the original stimulus, but need not correspond on a lexical or syntactical level.

The Reconstruction Hypothesis. James, et.al. look favorably upon these theories insofar as they concern translation of an objective stimulus into a coded idea in memory. Their own requirements for memory are sufficiently vague to be covered by either of the above theories: what goes into memory includes "a theme and basic logical relations [p. 53]." One more thing enters into memory. It is so obvious that it never seems worth mentioning: the knowledge that the stimulus was a grammatically complete sentence. What they are interested in is the process of retrieval of all of the information from memory. They are not concerned with revamping coding and storage theories except to the extent that this is a consequence of improved empirical testing following their appending a retrieval hypothesis to these theories.

The Reconstruction Hypothesis contains four fairly simple provisions. First, when the subject is reconstructing the stimulus sentence, he is likely to initiate it with the most salient noun, or the theme, which was recalled or abstracted from the original. Second, following James (1972), noun imagery is a compelling factor: high imagery nouns are more salient than low imagery nouns. Third, because the subject knows the stimulus was a complete sentence, he will consider the recalling of a complete sentence to be an implicit response demand. Fourth, because normal speech contains a bias toward
use of more active than passive voice sentences, this bias will be reflected in the reconstructed sentences.

**Voice and Theme.** The fourth of the premises of the Reconstruction Hypothesis requires some elaboration. Voice is an example of apparent redundancy of intent which seems to give the language user a choice of expressions for an idea. In a general sense, this is quite true; but there is good reason to believe that the decision between the two grammatical forms is not arbitrary. The simple, active declarative sentence has generally been the prototype of the basic form of organized information (Mehler, 1963). Based upon the assumption that speakers should produce more complex forms only for a purpose, the passive sentence can be seen as the result of a subtle change in focus.

A relevant distinction is between surface structure and logical structure. Surface structure refers to a grammatical organization of a sentence. The "subject" precedes the verb and the "object" follows the verb. The meaning of the sentence is ignored. Logical structure, on the other hand, refers to the meaningful relationships within the sentence. Each sentence (simple declarative sentence with a transitive verb) has an actor, which is the source of the activity of the verb, and a recipient of the action. Obviously, since grammatical structure is oblivious to content, the reversal of noun order following from a voice change causes the subject" to become the "object" and vice versa. Logical structure is governed by meaning and is not influenced in this way by a change in voice.

Because meaning (logical structure) is preserved despite a change in voice, the difference in focus mentioned earlier can be seen as a
matter of emphasis, the choice of a theme. If the speaker is more concerned with the actor than with the recipient, the sentence will be in active voice. If the recipient is more important, the passive voice will be used. With an idea quite compatible with the first provision of the Reconstruction Hypothesis, Clark (1965) and Halliday (1967) say that the subject tends to put the idea to be thematized at the beginning of the sentence. This is particularly true in the use of the passive voice which, being a more complex form than the active voice, implies some commitment to making a certain point on the part of the speaker.

The James, Thompson, and Baldwin (1973) experiment. A brief review of the experiment which tested the reconstruction hypothesis may be useful since it is the basis of the study reported here and because certain terms must be understood before predictions can be made. The stimuli in both experiments were simple active and passive sentences containing two nouns of rated imagery value, either high or low. They were constructed by orthogonal combination of imagery value and grammatical function: subject and direct object. This resulted in four sentence types for each voice. Using James, et al.'s coding system, which refers to the imagery of the nouns in their grammatical order, the four types are: HH, HL, LH, LL. (For example, HL refers to a sentence in which the first noun, or subject, is high imagery, the second noun, or direct object, is low imagery. This formula holds regardless of voice.) The task in the study by James, et al. was free recall of lists of such sentences; their analysis centered on the likelihood of retention or transformation of the voice of the
original stimulus.

The results of the James, et al. experiment were damaging to Paivio's theory and were interpreted to be generally supportive of Bransford, et al. if supplemented by the reconstruction hypothesis. The notion of imagery was not denied. Rather, imagery was assumed to determine saliency: the more concrete, the more memorable. The results showed that, in recalling sentences, subjects tended to initiate their responses with high imagery nouns. Thus, for example, active LH sentences tended to be transformed into passive HL sentences. Similarly, passive LH sentences were transformed into active HL sentences. The identification of saliency with memorability was further reinforced by a second study in the same paper which held noun imagery constant and through repetition of the "salient" noun, produced the same effect.

Criticism of the Reconstruction Hypothesis. Critics of imagery studies have complained that poor memory for meaning in abstract sentences reflects a problem of comprehensibility—the consequence of an inadequately controlled "meaningfulness" variable (Johnson, Bransford, Nyberg and Cleary, 1972; Klee and Eysenck, 1973)—despite Paivio's protestations to the contrary (see Paivio, 1969). The James, et al. experiment does not control for this factor either, which could seriously affect interpretation of their results both in the original coding phase and in the subsequent reconstruction of the sentences. The fact that one of those pointing out the problem, Bransford, is also one of the authors of the theory James, et al. endorse must at least make us question those results and their interpretation.
The obvious solution to the problem is to control noun selection for meaningfulness, as was done by Moesser (1974). Unfortunately, this is difficult to accomplish when a large number of sentences having several other requirements must be selected from a list of limited magnitude. Another solution is to use a paragraph context for the stimulus sentences to improve word interpretation. Memory for meaning of abstract sentences has been substantially improved by this procedure (Pezdek and Royer, 1974).

There are other reasons for using paragraph contexts to introduce the stimulus sentences. In the first place, it is doubtful that the constructive theory of Bransford and his colleagues can be fairly tested using individual sentences as stimuli. Most of the experiments on which their theory is based have used stimuli more complex than lists of words. In the individual sentence, the theme is so inextricably tied to an individual word that the influence of a gist, a general idea, may not be apparent. Use of entire paragraphs should allow differentiation between words and their psychological referents.

Another reason for the use of paragraphs is that imagery can be manipulated. Yuille and Paivio (1969) reported that memory for individual words was better when the stimulus paragraphs were concrete that when they were abstract. If this is true, the effect should be reflected in a better overall recall for sentences from high imagery paragraphs although the likelihood of voice transformation may not be affected. There is, however, reason to doubt Yuille and Paivio's findings. Morris and Reid (1972) attribute better recall to the imagery of the individually remembered nouns and not to the context
imagery. Because the nouns in the stimulus sentences are rated for imagery and cover both high and low imagery nouns, it should be quite simple to test that question within the study.
PRETESTING

The pretests were designed to determine how experimental subjects would perceive the stimulus materials. The individual nouns in the key sentences were drawn from a standard list which gave values to their imagery and meaningfulness. Three pretests were performed, two concerned with the contexts and one with the key sentences. There were four kinds of information sought:

1. Imagery values for the paragraphs rated as wholes (Pretests I and II);
2. Meaningfulness values for the paragraphs rated as wholes (Pretests I and II);
3. Meaningfulness values for the final key sentences rated in isolation and as part of the paragraph contexts (Pretest III);
4. Identification of which of the two nouns in the key sentence was closer to the theme of the context (Pretests I and II).

In the following sections, the construction of stimulus materials will be explained. Then each of the three pretests will be described. Finally, the results will be given within the framework of the four objectives just described.

Materials. Twenty-four simple transitive sentences were constructed using 48 nouns from Paivio, Yuille, and Madigan's (1968) list. Half of these nouns were considered to be high in imagery value (minimum imagery rating = 6.07) and half were of low imagery value (maximum imagery rating = 3.97). This differentiation was consistent with ratings from the same source used by many other experimenters (see, e.g., James, Thompson, and Baldwin, 1973; Groninger, 1974; Begg

11
and Paivio, 1969). Although an attempt was made to equate the two sets of nouns on meaningfulness and frequency, the concrete nouns were both more meaningful (mean ratings: high imagery = 6.54; low imagery = 4.96; Paivio, Yuille, and Madigan, 1969) and more common (mean occurrence: high imagery = 85.67; low imagery = 31.04; Kucera and Francis, 1968). The context was intended to reduce these differences.

The sentences were constructed by orthogonal combinations of noun imagery (high and low) and grammatical function (subject and direct object). The result was four sets of six sentences which contained only function words (subject, verb, direct object) and necessary articles. All were easily transformed in voice without distortion of basic meaning.

For each sentence, two paragraphs were constructed, one high and one low in overall imagery. Each paragraph was written to have its main theme most closely related to the logical-structure recipient-noun of the key sentence. The rationale for this manipulation will be discussed. The key sentence was always in the final position in the paragraph. None of the nouns in the key sentences appeared elsewhere in their passages.

Pretest I. Fifty-six paragraphs, including the 28 final key sentences (i.e., one high imagery context and one low imagery context per sentence) were divided into two lists of equal size. Each list contained 14 paragraphs written to be low in imagery value and 14 written to be high in imagery value. A paragraph for each key sentence appeared only once in either list. The key sentences were in opposite voices in the two lists, but otherwise choice of
voice and list was random. Eleven judges rated each list (22 in all).

Paragraphs were rated for imagery (7-point scale), meaningfulness (7-point scale), and theme. The thematic rating was actually a binary choice between the two nouns from the final sentence. Judges were asked to decide which of the two nouns from each key sentence was "most closely related to the theme, or the basic underlying, organizing idea of the paragraph." The paragraphs were tape-recorded by the experimenter. An attempt was made to speak in a flat, though not completely monotonous voice to avoid emphasis of particular words, sentences (including the key sentences), or paragraphs. A 30-second break between paragraphs allowed time for ratings.

Pretest II. This was identical to Pretest I except that the paragraphs were heard without final key sentences. There were several reasons for this difference. Since the paragraphs in this form were later to be used as cues whose imagery could influence recall, it was essential that those imagery ratings be known without the potential influence of the final sentence. By similar reasoning, the meaningfulness of the paragraphs had to remain at a reasonable level, although the lack of the final sentence, which usually finished a train of thought, was expected to reduce meaningfulness somewhat. The most important reason for Pretest II, however, was to confirm that the recipient noun of the last sentence was in fact the theme. This could only be determined with confidence if the ratings could be performed without the unknown influences of the final sentences. Fourteen judges performed the ratings, seven
for each group.

**Pretest III.** The last pretest was performed by 10 judges doing two rating tasks. They rated only meaningfulness (7-point scale). First they judged the 24 sentences selected in Pretests I and II for use in the experiment. The sentences, without contexts, were presented on tape with a 10-second interval between sentences to allow for rating. Again the presentation was in two lists equally divided for number of active and passive sentences, with each sentence appearing only once in a list. Between the two lists, each sentence was heard in both voices. Five judges heard each list.

Following this, the sentences were presented with their paragraph contexts and were rated again. Sentences were heard in the same voice as before and context imagery was randomly divided between lists. Instructions clearly specified that only the final sentence was to be rated.

**Context imagery.** The mean imagery values for the high and low imagery paragraphs finally used in the experiment are given in Table 1. The division into high and low imagery paragraphs was made on the basis of this information. The goal of the pretest was to assure that paragraphs could be constructed having different levels of abstractness while referring to the same key sentence. The mean difference between imagery ratings was 1.81 for Pretest I and 1.28 for Pretest II. The other important factor was that the imagery ratings not vary over too great a range across the various types of key sentences. The meaning of "too great" in this case is admittedly vague. A test of imagery ratings for Pretest I does
TABLE 1

Imagery Ratings (7-Point Scale) of the Paragraph Contexts with Key Sentences Present (Pretest I) and Absent (Pretest II).

<table>
<thead>
<tr>
<th>Sentence Type</th>
<th>High Imagery Context Pretest I</th>
<th>Low Imagery Context Pretest I</th>
<th>High Imagery Context Pretest II</th>
<th>Low Imagery Context Pretest II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active HH</td>
<td>6.07</td>
<td>4.02</td>
<td>5.75</td>
<td>3.85</td>
</tr>
<tr>
<td>Passive HH</td>
<td>6.65</td>
<td>5.83</td>
<td>6.50</td>
<td>5.00</td>
</tr>
<tr>
<td>Active HL</td>
<td>5.42</td>
<td>4.27</td>
<td>5.83</td>
<td>4.03</td>
</tr>
<tr>
<td>Passive HL</td>
<td>5.65</td>
<td>4.03</td>
<td>5.90</td>
<td>4.03</td>
</tr>
</tbody>
</table>

Mean 5.84 5.69 4.61 4.41
give some statistical support to the assumption that these goals were achieved. The contexts were seen as significantly different in imagery value \( F(1, 20) = 58.935; p < .001 \). No significant differences were found across sentence types. The differences between the pretests were small but clearly in the direction of lessening the distinctions.

**Context meaningfulness.** The average meaningfulness values for the high and low imagery paragraphs used in the study are given in Table 2. The purpose of this rating was to assure that the paragraphs made sense to the subjects. The results show a consistently high level of meaningfulness for Pretest I. The decline for scores in Pretest II reflects the fact that the paragraphs without the final sentences often failed to complete an entire idea and were, thus, seen as mildly ambiguous. In no case did a final sentence increase this confusion. The important result is that differences between high and low imagery paragraphs within a pretest are small enough that the contexts themselves create no recall variations independent of those resulting from imagery.

**Key sentence meaningfulness.** The results shown in Table 3 are the ratings of the final sentences only. They were presented to the judges either in isolation or at the end of a paragraph. As can be seen, the comprehensibility of the final sentences when they were part of a context varied according to the imagery make-up of the sentence. Lower imagery sentences were less meaningful than higher imagery sentences. Context imagery had little effect on the ratings. The use of a context, on the other hand, was clearly
TABLE 2

Meaningfulness Ratings (7-Point Scale) of the Paragraph Contexts with Key Sentences Present (Pretest I) and Absent (Pretest II).

<table>
<thead>
<tr>
<th>Sentence Type</th>
<th>High Imagery Context</th>
<th>Low Imagery Context</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pretest I</td>
<td>Pretest II</td>
</tr>
<tr>
<td>Active HH</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Passive HH</td>
<td>6.35</td>
<td>5.35</td>
</tr>
<tr>
<td>Active HL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Passive LH</td>
<td>6.14</td>
<td>4.91</td>
</tr>
<tr>
<td>Active LH</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Passive HL</td>
<td>5.75</td>
<td>4.65</td>
</tr>
<tr>
<td>Active LL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Passive LL</td>
<td>6.06</td>
<td>5.20</td>
</tr>
<tr>
<td>Mean</td>
<td>6.10</td>
<td>5.04</td>
</tr>
</tbody>
</table>
TABLE 3

Pretest III: Mean Ratings (7-Point Scale) of Key Sentence Meaningfulness

According to Sentence Type and Contextual Imagery

<table>
<thead>
<tr>
<th>Sentence Type</th>
<th>High Imagery Context</th>
<th>Low Imagery Context</th>
<th>No Context</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active HH</td>
<td>6.40</td>
<td>6.20</td>
<td>6.00</td>
</tr>
<tr>
<td>Passive HH</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Active HL</td>
<td>5.60</td>
<td>6.00</td>
<td>5.40</td>
</tr>
<tr>
<td>Passive LH</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Active LH</td>
<td>5.20</td>
<td>5.60</td>
<td>4.10</td>
</tr>
<tr>
<td>Passive HL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Active LL</td>
<td>5.20</td>
<td>5.20</td>
<td>3.20</td>
</tr>
<tr>
<td>Passive LL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>5.60</td>
<td>5.75</td>
<td>4.67</td>
</tr>
</tbody>
</table>
important. Higher imagery sentences were not much affected by the passages, but lower imagery sentences were dramatically influenced (e.g., 5.2 as opposed to 3.2 ratings for LL sentences with and without passages). These results were consistent with the predictions of Pezdek and Royer (1974) concerning the covariation of imagery and meaningfulness in sentences. The use of paragraphs evidently reduces the meaningfulness problem as was intended.

Choice of theme. One problem which arose with respect to the use of context did not have an entirely satisfactory solution. A coherent paragraph will have a theme which will have some relation to the sentences and nouns in the passage. This theme must be controlled, particularly when its effects are as important as they are here. Ideally, theme should have been varied over all the other central factors, but this proved to be too unwieldy for one study. Since the emphasis of the experiment is upon the effects of imagery, it was decided to attempt to stabilize the influence of theme as much as possible. This was accomplished by thematizing the logical recipient of the action in each of the simple sentence stimuli. Perfetti and Goldman (1974, 1975) found that, when the logical recipient (active-voice direct-object and passive-voice subject) is thematized and used as a retrieval prompt, the likelihood of syntactically accurate recall of active and passive sentences is approximately equal. Similarly, using the recipient as the theme makes the probability of voice transformation in recall about the same regardless of stimulus voice. When the logical agent is the theme, the probabilities for all
these conditions are quite divergent. Overall likelihood of transformation as compared to accurate recall is determined by many factors such as the length of the interval between stimulus presentation and recall, number of intervening items, and stimulus complexity. This solution, thematizing the logical recipient in the key sentences, while still not ideal, was considered the best option available. Such a method of determining theme, as opposed to focusing on the grammatical subject or object, has the additional advantage of permitting the use of the same contextual paragraphs for the same key sentences in both voices: logical structure is independent of voice while grammatical structure is not.

The proportions found in Table 4 indicate the likelihood of subjects choosing the logical recipient as the theme of the paragraph. The results, while in the expected direction, were not as clear as might have been desired. The paragraphs for the active LH/passive HL sentences stood out as being difficult to rate for theme. While only sentences whose ratings were in the direction of the recipient theme were used, the greater ambiguity of this particular group is disconcerting given the results of the study. It may be either that the sentences were poor stimuli which artificially produced voice changes or that they were in fact good stimuli and confusion comes as a natural consequence of trying to understand a sentence of this type.
TABLE 4

Proportion of the Ratings Indicating That the Logical Recipient of the Key Sentence Was the Theme of the Context.

<table>
<thead>
<tr>
<th>Sentence Type</th>
<th>High Imagery Context</th>
<th>Low Imagery Context</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pretest I</td>
<td>Pretest II</td>
</tr>
<tr>
<td>Active HH</td>
<td>.881</td>
<td>.926</td>
</tr>
<tr>
<td>Passive HH</td>
<td>.914</td>
<td>.863</td>
</tr>
<tr>
<td>Active HL</td>
<td>.794</td>
<td>.727</td>
</tr>
<tr>
<td>Passive LH</td>
<td>.828</td>
<td>.813</td>
</tr>
<tr>
<td>Mean</td>
<td>.854</td>
<td>.832</td>
</tr>
</tbody>
</table>
METHOD

Experimental stimulus organization. The 24 sentences and 48 paragraphs selected in pretesting were used as stimuli. Insofar as subjects are concerned, there was only one independent groups division: half heard all high imagery context and half all low imagery context paragraphs. The 24 paragraphs for each context condition were divided into two 12-item lists (hereafter designated A and B). A sham or unanalyzed paragraph was placed at the beginning of both lists to attenuate primacy effects. Three sentences from each sentence type appeared in each list. To offset order effects, all four sentence types (HH, HL, LH, LL) occurred once in each third of the list. Aside from these specifications, sentence placement was random. Voice types were also randomized between lists except for the limitation that half of the sentences of each list be in either voice. Counterbalancing was done across voice and list order. Half of the subjects heard List A and then List B; the other subjects heard them in the opposite order. For each list order, one group heard six active and six passive sentences per list; a second group heard those same sentences in the other voice. Thus a total of eight groups were created by orthogonal combination of context imagery (high and low), list order (A or B), and individual key sentence voice (active or passive).

The paragraphs were presented on a tape-recorder. In recording
the material, the same effort as described previously was made to avoid inflectional emphasis which might create coding or interpretation bias.

Subjects. Eighty unpaid volunteers from the summer school student body of the College of William and Mary were subjects. They were both men and women between the ages of 18 and 22. Testing was done in groups of one to four. Division into experimental groups was random. The results of nine other subjects were eliminated because of mechanical problems or very serious misunderstanding of instructions.

Procedure. The participants heard two sets of 13 paragraphs (12 test passages and the initial sham). Each paragraph was preceded by a bell-tone warning and its number in the order of presentation (e.g., "paragraph number one"). Following the reading of the paragraph was another bell-tone and a thirty second period during which the subject wrote the theme of the paragraph. The theme could be a word, a phrase, a sentence, or a paragraph. The intention of this activity was to get the subjects to listen to the entire paragraph and to encourage them to process it on a semantic level. After 30-seconds, the warning bell sounded again and the next paragraph was heard.

At the end of the list, the paragraphs were presented a second time, now without the final sentence. During this cued recall, the passages were in a different order and were designated by letters instead of numbers to avoid confusion. As before, a 30-second period followed each paragraph to be used for recalling the final sentence. There was a break of one minute between lists.
to allow the experimenter to change tapes. The second list was presented in precisely the same fashion as the first.

Pre-experimental instructions emphasized taking both writing the theme and remembering the final sentence with equal seriousness. Subjects knew they would later be trying to recall the final sentence of the paragraph and that they would be asked to do so as closely to verbatim as possible. A specific warning was made against paraphrase.

**Scoring.** The free recall data were evaluated by the experimenter following the guidelines contained in this section. No gradations in judgment were permitted: either a sentence achieved the predetermined criteria or it did not. Consequently, each recalled sentence yielded only a single categorical rating for each subject.

The most accurate level of whole sentence recall was termed Modified Verbatim. A category for strictly verbatim memory was rejected because, the verbs and articles not having been controlled, such recalled material is dependent upon unknown or poorly defined factors. A Modified Verbatim/Correct sentence was a simple sentence with the nouns written correctly and in the same sequence as in the original presentation. The voice of the verb was also identical to the original. Pluralization of a noun was considered accurate recall. Substitution of an incorrect verb or article or addition of an adjective were ignored. The truth value of the remembered sentence as compared to the original was not assessed. Thematic intrusions, such as addition of a prepositional phrase or a second clause, were allowed as long as they did not interfere
with the grammatical relationship of the two nouns.

The alternate form for this type of recall is Modified Verbatim/Transformed. The same rules as before apply except that the nouns are now in reversed sequence and the verb is in the opposite voice from the original.

The adept reader will have noted two important exclusions from the sentences allowed by the rules above. Consider the following sentences:

(A) The blasphemy was condemned by the priest.
(B) The blasphemy angered the priest.
(C) The priest was angered by the blasphemy.

Sentence (A) is one of the key sentences actually used in the experiment. Remember that the rules for Modified Verbatim recall allow a new verb to be substituted for the original without penalty. Thus, were that the only guideline, sentence (B) would be Modified Verbatim/Correct and (C), Modified Verbatim/Transformed. A closer look at sentences (B) and (C) will perhaps elucidate the problem. In (B), the grammatical-structure order of the nouns has been retained, but the logical-structure relationship has been reversed by the change in voice of the verb. In (C), the logical-structure relationship is upheld, but the grammatical structure is reversed. The addition of the rule concerning retention of voice and noun sequence for Modified Verbatim/Correct recall and the change of voice with noun sequence inversion for Modified Verbatim/Transformed recall rejects both (B) and (C). This result is neither arbitrary nor ideal. Consider another sentence:
(D) The blasphemy was applauded by the priest.

or, of course:

(E) The priest applauded the blasphemy.

Sentence (D) is an acceptable Modified Verbatim/Correct recall, and (E), a Modified Verbatim/Transformed recall; yet, both of these sentences are semantically inferior to (B) and (C). The rationale for the rules used is that (D) and (E) do not precipitate a clash between logical structure and grammatical structure, (B) and (C) demand a theoretical decision in favor of one or the other. An *a posteriori* excuse for the exclusion of sentence types (B) and (C) is the fact that they occurred infrequently. Analyses were performed using both interpretations (i.e., in one case interpreting a type (B) sentence as Correct and type (C) as a transformed sentence and vice versa). No significant change in the results ensued. Finally, it should be noted that, in the opinion of the experimenter—albeit a weak argument—no sentences so drastically at odds with the original meaning occurred as were found in examples (D) and (E). Such sentences would have been absurd when used to complete the contextuel paragraphs, which were used as response cues.

The existence of errors like (D) and (E) is by no means theoretically uninteresting. A study directed at discovering factors provoking one kind of error or the other may be useful in better defining the logical structure/grammatical structure difference. To the best of my knowledge, no such study exists.

A second level of recall was established to detect memory
less dependent on exact lexical memory. This category was called Partial Recall. Essentially the same rules as those for Modified Verbatim recall applied with the exception that only one noun needed to be remembered. Using the above example, a sentence could be:

(F) The blasphemy was condemned by the minister.

or,

(G) The blasphemy was condemned.

A Partial Recall/Transformed sentence similarly follows the Modified Verbatim/Transformed rules. Again, the truth value of the recall was not assessed.

For the purpose of sentence analysis, Partial Recall was not considered a separate process. Although the act of retrieval leading to Partial Recall is undoubtedly different from that leading to Modified Verbatim recall, the experiment was not designed to ferret out this distinction. Partial Recall was introduced in recognition of the fact that the Modified Verbatim category eliminated a great many instances of the reconstruction process. Thus, for analysis of sentences, the data from the two categories were combined. As a matter of convenience, the combined Partial Recall data and Modified Verbatim data will be referred to as Total/Voice Correct and Total/Voice Transformed. As these titles imply, the assumption is given that syntactic memory is independent of the sort of verbatim memory detected by the Modified Verbatim rules.

Some final notes on procedure. A discussion of one major technical
problem has been delayed until now so that the reader would have sufficient information to understand a departure from accepted organization.

The sentences heard by each subject are divided into two lists, A and B, consisting of 12 passages and a sham or distractor. As a result, some learning effects might be expected, both within and between lists. Several things were done to minimize these effects: (1) one example of each sentence type appeared in every third of the list (i.e., the lists were divided into three sets of four sentences); (2) lists were counterbalanced in order of presentation between groups of subjects; (3) subjects were informed of the recall task prior to the start of the experiment. This final point is the most difficult to justify; the rationale for the others is rather self-explanatory.

The purpose of having informed subjects was to equalize expectations between lists. Certainly it would be ridiculous to surprize a person with a recall task following list A and then expect this person to revert to a naive state to be surprized again after list B. The danger, of course, is that the subjects will spend more time memorizing key sentences than listening to the entire passages. Two methods were used to offset memorization: (1) writing of the theme of the paragraph between each presentation in the learning phase; (2) reordering of the paragraph cues in the recall phase and informing the subjects of this fact beforehand. There is no doubt that most people spent more time rehearsing the key sentences than they would have using the
surprise-recall tactic. Hopefully, this does not negate the validity of the experiment; it merely influences interpretation of the results.

The most basic question is: why divide the stimuli into two lists at all? Several factors combined to make listing the least undesirable procedure. A minimum data requirement was the overriding reason. Each subject, as mentioned earlier, contributed only 24 units of categorical data, these divided among several variable combinations. To any one cell, a subject could contribute only one categorical response. The obvious remedy is increase the number of stimuli each person receives. However, the experiment as limited to the 24 passages took approximately an hour to complete. The methodology, therefore, creates its own quantitative limits. The alternative is to run more subjects. Unfortunately, the available pool was almost entirely exhausted.

All this explains the 24-items, but not the use of the two lists. A precedent for the 12-item paragraph lists was set by Perfetti and Goldman (1974). In their study, however, only one list was required to derive adequate information. Presentation of 24 passages seemed an inordinate amount of interference prior to recall. When this was attempted in an earlier heuristic study, much more drastic order effects did seem to emerge than were evident in the arrangement used here. No reliable statistical evidence can be offered to support this statement, so it is possible that the cutoff used was, in fact, rather arbitrary. The
best explanation that can be given is that, after attempting various list lengths, the 12-item list appeared to allow sufficient interpolated material to reduce memorization effects (i.e., verbatim recall of about 50%) while not being prohibitive of moderately precise memory.
RESULTS AND DISCUSSION

Organization of the data. An article by Clark (1973) on the proper analysis of language data has contended that such materials are frequently incorrectly characterized as fixed-effects variables. They are statistically treated as an entire population, but are then interpreted as samples of a larger population. To avoid this error, both subjects and Sentence Types (HH, HL, LH, LL) should be treated as random effects. Unfortunately, this is practically impossible because, as noted before, each subject contributes so little to any one cell. To avoid too much confusion, collapsing across either sentences or subjects seemed desirable. The more traditional organization of the data calls for averaging or totaling across the six sentence responses, providing ten scores (one for each subject) as a source of variance. However, as was explained in the Method section, the only between-groups variable was Context Imagery. Within each sentence type, three of the six sentences were in active voice and three in passive voice. For counterbalancing purposes, the three sentences in each voice were divided between lists, one or two in each. In other words, each subject was involved in so many different factors as to make it difficult to use the Subjects variable in the usual way.

It was decided to collapse across subjects and to use the number of responses within a particular scoring category as data. This permits a score of zero to ten (ten subjects responded to each
sentence under any one variable combination) for each sentence, with six scores per cell. With sentences rather than subjects as reference, Sentence Type becomes the only between-groups variable. Voice (active and passive), Contextual Imagery (high and low), and List (first or second in order of presentation) become within-sentence variables. Thus main analyses used a four factor analysis of variance with repeated measures on 3 factors: Sentence Type x Stimulus Voice x Contextual Imagery x List. For analyses combining both Correct and Transformed data, the same design was used; this being accomplished by averaging across lists and eliminating the List factor. In the place of that factor the Response Voice was added, also a repeated factor.

The way the data were analyzed will not be directly reflected in much of the following discussion. This is because the discussion is centered around a grammatical organization of the data in which an active and a passive sentence of a certain Sentence Type do not necessarily have the same content; they correspond only in their grammatical structure related to noun imagery. For example, take the following sentence:

(G) The student sought truth.

This is an active HL sentence ("student;" high imagery grammatical subject; "truth;" low imagery grammatical object). A grammatical equivalent of (G) in the passive voice is:

(H) The soldier was condemned for treason.

This is a passive HL sentence ("soldier;" high imagery grammatical subject; "treason," low imagery grammatical object).
This kind of internal structure of the sentence is of interest to James, Thompson, and Baldwin because they are interested in response tendencies reflected in temporal or spatial priority in recall of the nouns in the sentence. While this is a reasonable theoretical focus, use of sentences such as (G) and (H) as basically equivalent except for voice ignores the impact of meaning — of logical structure. Studies which equate sentences of the same grammatical structure, e.g., HL, across voices makes the very assumption they intend to test: that a sentence of a particular grammatical structure is involved in the same coding and retrieval dynamics as is any other sentence of that grammatical structure, regardless of voice. Such an assumption is not justified. Meaning is very likely not invariant across voices (see e.g., Tannenbaum and Williams, 1968; Olson and Filby, 1972), but it has great stability, the variation being in emphasis.

This reasoning is behind the use of sentences as a random-effects variable. The other point to be made, however, is that, for purposes of analysis, the passive "equivalent" of (G) will not be (H) but the following:

(I) Truth was sought by the student.

This is a passive LH sentence ("Truth," low imagery grammatical subject; "student," high imagery grammatical object). This makes analysis of differences among Sentence Types a trickier activity than it might otherwise have been. Such pairing does succeed in eliminating uncontrolled variation in meanings which could produce the appearance of voice effects which are not real. The data is transposed to a grammatical structure organization — i.e., pairing Sentence Types (G)
and (H)—for discussion. This involves no change of the results, but it does require care in analyzing relationships among groups and conditions.

**Key Sentence Recall.** The overall recall of the stimuli was clearly a function of the stimulus sentences. Table 5 shows the percentage of the sentences recalled from a possible 240 in each voice, or 480 total for each sentence type. Sentences are grouped here according to their logical structure, which is to say according to content. The consistency of recall between voices is so obvious that there is no doubt that the simple fact of memorability was unaffected by stimulus voice. This consistency also held up through a change in scoring criteria, i.e., for Partial Recall category.

The level of recall is quite high in comparison with the similar recall task of James, et.al. who, for their most liberal scoring category, showed a variation between 46% recall of HH sentences and only 7% recall of LL sentences. Their results also showed no effect of stimulus voice in recall. The variation in recall according to sentence type in the present study is centered around the active LH/passive HL sentences. Only about half as many of these sentences were recalled as were any other type. As will be discussed in a later section, the distinctiveness of this sentence type leads to the possibility that an entirely different kind of sentence processing is occurring in this case than for other types.

**Recall in the Original Voice.** The main effect for Stimulus Voice was the only factor or interaction to achieve or even approach significance in the Modified Verbatim/Correct \[F(1,20) = 9.764; p < .01\]
TABLE 5
Percent of the Stimulus Sentences Recalled,
Combining the Correct and Transformed Sentences.*

<table>
<thead>
<tr>
<th>Stimulus Voice</th>
<th>Active (240)</th>
<th>Passive (240)</th>
<th>Total (480)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active HH</td>
<td>57% (88%)</td>
<td>50% (88%)</td>
<td>54% (86%)</td>
</tr>
<tr>
<td>Passive HH</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Active HL</td>
<td>50% (78%)</td>
<td>49% (78%)</td>
<td>50% (78%)</td>
</tr>
<tr>
<td>Passive LH</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Active LH</td>
<td>26% (63%)</td>
<td>20% (61%)</td>
<td>23% (62%)</td>
</tr>
<tr>
<td>Passive HL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Active LL</td>
<td>44% (75%)</td>
<td>44% (80%)</td>
<td>44% (78%)</td>
</tr>
<tr>
<td>Passive LL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>44% (76%)</td>
<td>41% (77%)</td>
<td>43% (76%)</td>
</tr>
</tbody>
</table>

* In each cell, the first number is from the Modified Verbatim scoring category. The parenthesized number is from the Partial Recall data.
and the Partial Recall/Correct $[F(1, 20) = 5.657; p < .05]$ analyses. The cause of the smaller F-ratio in the second analysis is not a decline in the difference between Stimulus Voice, but rather an increase in Voice x Sentences—within-groups error, which is not surprising given the freer scoring criteria for Partial Recall. The significant effect means that active voice sentences were recalled correctly more often than were passive voice sentences, regardless of Sentence Type.

This particular result is not predicted by any of the theories cited earlier. The Reconstruction Hypothesis does call for a response preference for an active over passive voice, but it could hardly claim to have predicted or explained the better stimulus qualities of an active voice sentence. The other theories do not even do this well.

Table 6 shows the total number of sentences correctly remembered in each voice condition. In reading this table and Table 7, it should be remembered that these sets of data are not independent. For example, the sentences under "Active" in Table 6 and those under "Active to Passive" in Table 7 are different responses to the same stimuli. So, of 240 active HH sentences, 115 were correctly recalled and 21 were transformed under the Modified Verbatim rules. This division of the data is somewhat confusing, but it succeeds in isolating the different recall processes better than does lumping all into one table.

Table 6 shows that there was some variability among groups in correct recall. Any possible effect was eliminated by a high between-sentences and sentences—within-groups variability. For the Voice effect, the only contradictory case is for LH sentences in the Modi-
TABLE 6
Total Number of Key Sentences Recalled
In Original Stimulus Voice. *

<table>
<thead>
<tr>
<th>Sentence Type</th>
<th>Active</th>
<th>Passive</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>HH</td>
<td>115 (164)</td>
<td>81 (141)</td>
<td>196 (305)</td>
</tr>
<tr>
<td>HL</td>
<td>110 (169)</td>
<td>41 (116)</td>
<td>151 (285)</td>
</tr>
<tr>
<td>LH</td>
<td>51 (114)</td>
<td>49 (76)</td>
<td>100 (190)</td>
</tr>
<tr>
<td>LL</td>
<td>99 (157)</td>
<td>78 (137)</td>
<td>177 (294)</td>
</tr>
</tbody>
</table>

Total 375 (604) 249 (470) 624 (1074)

* In each cell, the first number is from the Modified Verbatim scoring category. The parenthesized number is from the Partial Recall data. Total possible in each cell is 240.
TABLE 7

Total Number of Key Sentences
Transformed in Voice.*

<table>
<thead>
<tr>
<th>Sentence Type</th>
<th>Active to Passive</th>
<th>Passive to Active</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>HH</td>
<td>21 (48)</td>
<td>40 (71)</td>
<td>61 (119)</td>
</tr>
<tr>
<td>HL</td>
<td>11 (17)</td>
<td>6 (30)</td>
<td>17 (47)</td>
</tr>
<tr>
<td>LH</td>
<td>12 (38)</td>
<td>69 (112)</td>
<td>81 (150)</td>
</tr>
<tr>
<td>LL</td>
<td>6 (24)</td>
<td>27 (54)</td>
<td>33 (78)</td>
</tr>
</tbody>
</table>

Total 50 (127) 142 (267) 192 (394)

* Table 7 is organized exactly like Table 6.
fied Verbatim analysis, 425 active stimuli and 391 passive stimuli were recalled. The overall recall of Partial Recall sentences was even closer between voices: 731 active stimuli and 737 passive stimuli. An analysis of all the Modified Verbatim data (i.e., a combination of Modified Verbatim/Correct and Transformed) showed the voice of the stimulus to be a nonsignificant main effect.

Passive to active voice transformation was clearly the more common occurrence. Approximately one-third of the passive stimuli recalled were transformed to the active voice (36.32% for Modified Verbatim/data; 36.23% for Partial Recall data). The Sentence Type with the greatest likelihood of transformation was the passive LH sentences. HL sentences, on the other hand, were quite resistant to transformation.

Table 8 provides a useful view of the results as well as a comparison of these findings with those of James, et al. The probabilities for the James, et al. study are taken directly from their 1973 paper [p. 57]. The method for computing probabilities is theirs (total number of sentences transformed from a particular voice divided by the total number of stimulus sentences in that voice recalled either correctly or transformed). Partial Recall data from the present experiment were used because the scoring criteria for that category more closely approximated James, et al.'s standards that did the Modified Verbatim rules. The numerical values of these probabilities cannot be fairly compared between experiments because of the major methodological differences. The relative magnitudes can, however, be compared; they form the core of the conclusions.
### TABLE 8

Transformational Probabilities of Key Sentences Comparing
James, Thompson, and Baldwin (1973) and the Present Experiment.

<table>
<thead>
<tr>
<th>Sentence Type</th>
<th>Active to Passive</th>
<th>Passive to Active</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>James, et. al.</td>
<td>Carroll</td>
</tr>
<tr>
<td>HH</td>
<td>.053</td>
<td>.226</td>
</tr>
<tr>
<td>HL</td>
<td>.029</td>
<td>.091</td>
</tr>
<tr>
<td>LH</td>
<td>.245</td>
<td>.250</td>
</tr>
<tr>
<td>LL</td>
<td>.070</td>
<td>.133</td>
</tr>
<tr>
<td></td>
<td>.174</td>
<td>.335</td>
</tr>
<tr>
<td></td>
<td>.049</td>
<td>.206</td>
</tr>
<tr>
<td></td>
<td>.333</td>
<td>.596</td>
</tr>
<tr>
<td></td>
<td>.163</td>
<td>.283</td>
</tr>
</tbody>
</table>
The consistencies across experiments are obvious: LH sentences of either voice are highly susceptible of transformation. Active HL sentences are very resistant to transformation. The same can be said of passive HL sentences, although the degree of difference between passive HL and other passive sentence types is not so great as James found. Passive and active LL stimuli and passive HH stimuli showed an intermediate transformational probability between those of HL and LH sentences. This accords well with James' expectations. Active HH sentences were transformed to passive voice relatively more often than occurred for James, et.al. Active HH and LH sentences had very similar transformational probabilities.

These results actually conform in many elements with the predictions of the James, et.al. study. The relative imagery values of the two nouns in the stimulus sentences is the determinant of probability of transformation: if the first noun is higher in imagery than the second, the sentence is very resistant to transformation. When the first noun is of a lower imagery value than the second, the sentence is more susceptible of transformation. When the nouns are of equivalent imagery value, the probability is intermediate of these two forms. This is an accurate description of the predictions of James' reconstruction hypothesis. The greater stability of active voice, another of the predictions of the reconstruction hypothesis, is also evident in the results.

The one non-conforming feature of these results is the greater likelihood of transformation for HH than for LL sentences in both stimulus voices. This difference is not statistically significant.
according to Tukey's HSD, but then the only significant HSD is between passive LH and HL sentences [HSD(q.05; 4,20)= 2.56; $X_{HL} - X_{LH} = 3.42$].

Other key sentence variables. Some mention should be made here of those factors which were tested and found not to have produced any significant effects. The most important of these was the imagery of the context, for which the F-ratios never even approached significance. The paragraphs were certainly not ignored; they were the recall cues. Pretesting indicated that the contexts also had the effect of reducing meaningfulness differences among sentences of various imagery values. There is good reason to suppose that this meaningfulness equalizing was also occurring in the experiment. James, et al. found overall recall of sentences to vary according to Paivio's ordering of sentences: HH, HL, LH, LL (in decreasing order of memorability). The effect was very obvious, with approximately four times as many HH sentences remembered as were LL sentences. The variation in memorability among sentence types in the present experiment was quite small by comparison: the largest difference here—between HH and HL types—was slightly less that the smallest difference—coincidentally for the same two types—in James' study. Such an improvement in memorability was predicted by the author of the present experiment as a consequence of improved meaningfulness of low imagery nouns produced by the use of a context.

The other main factor, which was a methodological check on list effects, also produced no significant results and no indication of a trend toward significance. This merely indicates that any practice effects from the first list to the second were negligible.
Neither of the above two variables was involved in any significant interactions.

**Lexical memory.** The primary reason for investigating word memory was to find the effect of the imagery of the context upon memory for items of high and low imagery value. No significant effects of Contextual Imagery on memory for words was found. This result, along with the lack of effect of Contextual Imagery on sentence memory, argues strongly against Paivio's contention that a high imagery paragraph enhances memory for items contained within the paragraph.

Several other interesting statistics emerged from the word analyses. Table 9 gives the total number of nouns recalled by Voice, function and Sentence Type from a possible 240 in each cell. A significant Voice effect for grammatical subjects \( F(1,20)=11.163; p<.01 \) and for grammatical objects \( F(1,20)=11.186; p<.01 \) clearly indicates that the subject, or the first word, of an active sentence and the object, or second word, of a passive sentence are the better remembered of the two nouns in each sentence. The greater tendency for the passive to active voice transformation over the active to passive direction is paralleled by this greater recall of the passive object-noun than subject-noun. Similarly, the relative stability of the active is reflected in the subject recall.

There were also two significant Sentence Types effects for both subject-nouns \( F(3,20)=4.588; p<.05 \) and object-nouns \( F(3,20)=3.597; p<.05 \). This seems primarily the result of the fact that nouns from the LH sentences were, regardless of imagery value or logical or grammatical function, less memorable than nouns of other Sentence Types.
TABLE 9

Total Number of Nouns Recalled.*

<table>
<thead>
<tr>
<th>Sentence Type</th>
<th>Active Stimulus</th>
<th>Passive Stimulus</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Agent Subject</td>
<td>Recipient Object</td>
<td>Agent Object</td>
</tr>
<tr>
<td>HH</td>
<td>210 (H)</td>
<td>154 (H)</td>
<td>203 (H)</td>
</tr>
<tr>
<td>HL</td>
<td>188 (H)</td>
<td>144 (L)</td>
<td>190 (L)</td>
</tr>
<tr>
<td>LH</td>
<td>133 (L)</td>
<td>94 (H)</td>
<td>124 (H)</td>
</tr>
<tr>
<td>LL</td>
<td>161 (L)</td>
<td>134 (L)</td>
<td>173 (L)</td>
</tr>
<tr>
<td>Total</td>
<td>692</td>
<td>526</td>
<td>690</td>
</tr>
</tbody>
</table>

*Parenthesized letters indicate the noun's imagery value.

Total possible recall in each cell is 240.
Finally, it should be mentioned that high imagery nouns were, as individual stimulus items, only slightly more memorable than low imagery items. 1256 high imagery nouns and 1147 low imagery nouns were recalled, a difference of slight dimensions given the importance attributed to imagery by Paivio.
CONCLUSIONS

In general, the predictions of James, Thompson, and Baldwin (1973), combining Bransford's "construction hypothesis" with their own "reconstruction hypothesis," held up well in an experimental situation quite different from their own. Of primary interest to their theory are the transformational probabilities of the different Sentence Types. Entirely new stimulus sentences were written; recall was cued, not free; stimulus presentation was auditory, not visual. These are a few of the methodological differences between the studies; yet the order of voice transformation predicted by James, et al., was upheld.

The hypothesis which combines the theories of Bransford (constructive memory) and James (reconstructive recall) predicts most of the results. This hypothesis says that information as stored in memory is thematic and non-syntactical. In recall, the most salient noun in memory is retrieved first and is used as the basis for the reconstruction of the sentence. Higher imagery items tend to be more salient and are thus more likely to be recalled first. Since saliency is determined by relative imagery, LH sentences should frequently be turned into HL sentences. Of intermediate transformational probability should be HH and LL sentences, which have no relatively more salient noun. The most stable sentence type should be HL sentences, whose high imagery noun is already first.

The study reported here was intended to question the validity
of the concept of imagery, suggesting that imagery values were in practice confounded by noun meaningfulness differences between imagery levels. To some extent, these criticisms were justified by the results. Differences in memorability among various Sentence Types were drastically reduced in comparison to the James, et.al. study, and this despite the lack of full success in matching sentence meaningfulness across Sentence Types in pretesting. Similarly, the imagery of individual nouns in the sentences was of little consequence in their likelihood of recall. The results suggest that memorability of words or sentences need not be influenced by imagery value.

Still, the effects of imagery were not totally eliminated as predicted by the author. Relative imagery value was the basis of the dynamics of voice transformation described above. The problem now, however, is with the meaning of the term "saliency". High imagery items were only a bit more likely to be recalled than low imagery items, certainly not enough to produce the results found here. Table 7 confirms the yet more confusing fact that individual noun memorability organized by grammatical function or priority in recall, while it may explain the relative instability of the passive voice, does nothing to explain differing transformational probabilities among sentence types. The same table shows that the theme of the paragraph (and thus of the recall cue), the logical recipient as determined in pretesting, was not the force underlying voice transformation. The results apparently argue for the view that imagery determines saliency; but saliency is something other than memorability. James, et.al. suggest vividness as the defining mental experience,
but their obvious intention is to state in another, more experien-
tial way their finding of better recall for the salient, high imagery
word.

Bransford's constructive hypothesis certainly does not explain
the orderly behavior of voice transformation nor does it give any
hint of a definition of saliency. The addition of the reconstruc-
tive hypothesis is no improvement. Paivio states that subjects tend
to use high imagery items as recall cues whenever they can. Unfor-
tunately, Paivio's reasoning is so tied to vividness of high imagery
nouns, and vividness defined by memorability, that the notions can
hardly be separated. In other words, the several theories being
tested really shed little light on the concept of saliency when it
is stripped of its memorial function.

The experience of the author in writing the stimuli did suggest
an avenue of investigation which is, in spirit if not in fact, an
extension of the reconstruction hypothesis. Simple LH transitive
sentences are difficult to write. Because all the sentences were
in fact written first in the active voice, the experience was one of
trying to attribute agency to an abstract notion whose effects were
directed at a concrete recipient.

Hypothesis, induction, deriving the general from the effects
of the particulars is a characteristic mode of modern thought. The
ideas of universality and abstraction are capable of having the same
referent, but the universal is an archaic concept and abstraction con-
tains the strong suggestion of being a derivative process. We ab-
stract from the facts to produce a general idea. Hypostasis, de-
duction, the dependency of the particular on the general is difficult for those not well trained in mathematical thinking. Even for such people, it is probable that deduction is more a tool than a preferred conceptual category.

The suggestion being made is that the difference in transformational probability among the sentence types is, as James suggests, caused by response tendencies. These tendencies are the result of certain characteristic ways of thinking, not upon the storage experience-mode of imagery. The abstract/concrete distinction has more cognitive importance than the high/low imagery distinction. Abstracts interact easily with abstracts—Beauty and Justice are near neighbors in the hierarchy of Platonic Forms. Similarly, in the world of particulars things easily interact, for such is our daily perceptual experience. But when the abstract and the concrete must interact, our familiar ways of thinking take over. We need a concrete idea to factually justify the existence of an abstract idea. Saliency in this sense is akin to Paivio's "conceptual-peg", not as a recall cue but as a conceptual focus. The concrete idea is the logical center of our thinking, which implies nothing about relative memorability.

The implication of all this is that people find active LH sentences awkward. e.g.:

(J) The custom confused the gentleman.

The passive version logically pegs the human being as the focal element. The passive LH sentence is both unfamiliar and periphrastic, e.g.:
Obedience was demanded by the king. The impulse here is to simplify and clarify.

This proposal is obviously highly speculative, it was not tested in this study and, therefore, is not put forth as the acceptable explanation of the results. What is suggested is that none of the theories being tested adequately handles the results. The combination of Bransford's and James' theories is promising, but the allegiance of James to Paivio's imagery theory is not justified. Bransford, Barclay, and Franks (1972) speak directly of the concept of imagery, finding it inadequate by being too closely tied to the specific linguistic input. The addition of the reconstruction hypothesis with its base in imagery, to their constructive hypothesis is, consequently a dubious undertaking. The new proposal found here is an attempt to reconcile the predictable results of James' reconstruction process with the requirements of Bransford's constructive memory.
APPENDIX

Experimental stimuli: Paragraphs and key sentences.

Paragraphs are arranged according to the imagery of the final sentence: HH, HL, LH, LL. The high imagery paragraph is always followed by an active voice sentence and the low imagery context by a passive voice sentence. In the study, both voices were used with both types of contexts. Also included are the rated imagery (7-point scale) of each paragraph (Pretest I) and the percentage of pretest judges who rated the recipient noun of the key sentence to be the theme of the paragraph.
Just as she began to fall asleep, Mrs. Grimms was awakened by the loud celebration in a nearby apartment. Her head throbbed painfully and her back was hurting. The old woman clenched her fists as she heard her daughter's child, who had been sleeping in the next room, begin to cry.

THE PARTY IRRITATED THE GRANDMOTHER. (HH)

For a long time, Mrs. Grimms had persevered despite other people's thoughtlessness. But, just as age may cause resignation, it can also reduce one's tolerance. And the old lady's concern for her daughter's child, sleeping next door, made her attitude even less amiable.

THE GRANDMOTHER WAS IRRITATED BY THE PARTY.

The huge grey steeple of the old Gothic cathedral stood threateningly over the battlefield. But even its massive stone walls could not withstand the fury of the battle.

A MISSLE RUINED THE CHURCH. (HH)

The old Gothic structure long stood as a symbol of the grandeur and hope of the Middle Ages. In the end, it disappeared in a war which ignored such symbols.

THE CHURCH WAS RUINED BY A MISSLE.
The Roman town of Pompeii was once a bustling marketplace. Merchants, craftsmen and screaming children filled its cobblestone streets. It contained many temples and luxurious mansions.

A VOLCANO DESTROYED THE CITY. (HH)

Modern classical scholars and archeologists have found the Roman town of Pompeii a useful source of historical information. Prior to its demise, Pompeii was a major administrative center for southern Italy.

THE CITY WAS DESTROYED BY A VOLCANO.

The accused man stood with his head held high and his eyes staring resolutely at his accusers. Loving other men more than stupid laws, he had dared to steal from the rich merchant to save the life of a starving.

THE JURY CONVICTED THE PRISONER. (HH)

The accused man refused to believe that the law existed as a good in itself or was not superceded by the importance of the individual. Placing his fellow man above the law, he had helped others while ignoring bureaucratic regulations.

THE PRISONER WAS CONVICTED BY THE JURY.
The jewelry dealer brought out a black trunk and opened it to reveal a shining collection of expensive-looking gems. Several people gasped in admiration, but one man looked closely at the stones. He picked up one of the jewels and tested it in a surprising way.

A HAMMER CRUSHED THE DIAMOND. (HH)

The art of counterfeiting goes beyond the reproduction of government issued currencies and includes, among other things, jewelry-making. In a recent case, the authenticity of a Russian religious icon was called into question. One of its ornaments was removed and the gem was subjected to a simple test.

THE DIAMOND WAS CRUSHED BY A HAMMER.

The bullet wound was bleeding badly when the physician arrived. A dark red stream poured from the hole and the rug beneath the man was deeply stained.

THE DOCTOR STOPPED THE BLOOD. (HH)

Having found that the plasma was infected, the physician immediately notified the company which was transporting it to the hospital. The effects of a transfusion of the diseased substance on a weakened patient could be serious.

THE BLOOD WAS STOPPED BY THE DOCTOR.
The bearded young man stood before the sullen minister and loudly proclaimed that God did not exist. For a moment, there was a shocked silence throughout the church. Then the angry people in the congregation rose with a deafening cry of: "sacrilege."

THE PRIEST CONDEMNED THE BLASPHEMY. (HL)

The writings of the young theology student were brought to the attention of the church leaders. The papers supported many censored scientific notions. The materials were studied by a religious authority.

THE BLASPHEMY WAS CONDEMNED BY THE PRIEST.

The officer saluted and quickly turned to carry out the order. Without a question, the troops lined up and stood so rigidly at attention that they looked like a row of stone men.

THE KING DEMANDED OBEDIENCE. (HL)

The royal army was well known for its discipline and faithfulness. Every order was followed immediately and without question.

OBEDIENCE WAS DEMANDED BY THE KING.
The weeping peasants were lined up against a wall and, with a loud blast of the rifles, were mowed down by the firing squad. The general attempted to hide the bloody killing of innocent people, but a photographer, hidden in a nearby building, took many pictures of the horrible murder.

A NEWSPAPER REPORTED THE ATROCITY. (HL)

Violent actions by federal agencies are typically subject to tight security, but, in this case, the irresponsibility was revealed. Innocent people had been slaughtered without provocation.

THE ATROCITY WAS REPORTED BY A NEWSPAPER.

The young general stood sternly before his soldiers and screamed that they could never be defeated. With a cheer of hopeless rage, the men charged the huge stone fortress.

THE ARMY BELIEVED THE MISCONCEPTION. (HL)

The Napoleonic legions were firmly convinced that they were invincible. The mistaken concept was widespread throughout France.

THE MISCONCEPTION WAS BELIEVED BY THE ARMY.
The woman had long ago forgotten her happy childhood when she had played with her friends among the trees and gardens of her home. Now she was old and crippled and, until one day rumaging through her old trunk, she had not tried to look back on those warm summer days and her carefree laughter.

A DOLL REVIVED HER MEMORY. (HL)

Often we do not remember past events until some previously common object is again witnessed and serves as a stimulus to an entire chain of reflection. One amnesia patient had completely blocked out her early life until she saw something she had once found very important.

HER MEMORY WAS REVIVED BY A DOLL.

The mood in the auditorium was depressed and uncomfortable as the seemingly interminable recitals dragged on. The incompetence of the musicians was bad enough, but the poor acoustics and the inadequate ventilation made the situation nearly intolerable. Until the last recital, everyone had thought the evening a hopeless waste.

THE PIANIST DISSIPATED THE BOREDOM. (HL)

People yawned openly as they watched the young musicians fumble embarrassingly with their instruments, making the music sound flat and ridiculous. Everyone was tired and unhappy. But, when the last young man began to play, the complaining stopped.

THE BOREDOM WAS DISSIPATED BY THE PIANIST.
The magistrate's face reddened as he listened to the defendant's bitter sarcasm. Although he was usually very dignified and composed, he was noe becoming irritated. Rising, the black-robed figure demanded order in the courtroom.

THE ANIMOSITY DISPLEASED THE JUDGE. (LH)

The magistrate, an eminent and respected legal expert, was somewhat disconcerted by the controversy that prevailed in his courtroom. In his thirty years on the bench, his authority had never been questioned. He felt compelled to assert his authority before a travesty was made of the judicial process.

THE JUDGE WAS DISPLEASED BY THE ANIMOSITY.

Colonel Martin looked out the window at the demonstrators. His hand trembled as he tried to restrain his impulse to rush out and thrash them.

PACIFISM INFURIATED THE OFFICER. (LH)

Colonel Martin contemplated the accusations made against his theory of national defense. His conduct and beliefs had always been based on patriotism and manly integrity, and not, as others suggested, on the egotism of a frustrated tyrant.

THE OFFICER WAS INFURIATED BY PACIFISM.
Much of the grain still lay unbundled on the ground as the black storm clouds darkened the field. If it got wet, the precious crop would rot and be ruined. The farmer had nearly given up hope of saving it when a truck full of his friends appeared on the road.

COOPERATION SAVED THE WHEAT. (LH)

Grain crops had long been the primary source of income for the valley. The prediction of an early frost, which would limit time available for harvesting, now demanded unified action. All the men knew that destruction of the precious grain would also end the prosperity to which they had grown accustomed.

THE WHEAT WAS SAVED BY COOPERATION.

Sir Thomas Markham, the famous explorer, had travelled far to study the small African tribe. He had climbed tall mountains and gone through jungles and dangerous rivers to reach this hidden civilization. When he arrived in the village, however, he was immediately disrobed and brought before the chief.

THE CUSTOM CONFUSED THE GENTLEMAN. (LH)

Sir Thomas Markham, the famous Anglo-Saxon scholar, had studied intensively the newly-discovered ancient writings. As the foremost expert in his field, he found most of the translation quite easy. However, one recurrent symbolic prose form continued to evade his comprehension.

THE GENTLEMAN WAS CONFUSED BY THE CUSTOM.
The young man sat quietly at his desk and opened his notebook. He looked without interest at the lecturer and was prepared to take notes. But, as the professor began to speak, he found he was unable to write or concentrate and he shook with frustration over this meaningless education.

THIS NONSENSE ANGERED THE STUDENT. (LH)

The young man had found that the academic types of information he learned in college led only to stagnation and boredom. He was bright and worked hard, but was frustrated by the feeling he was wasting his time.

THE STUDENT WAS ANGERED BY THIS NONSENSE.

Once again, the old man stood before the dull, uninterested faces of his students. As he became involved in his talk, however, he spoke with excitement, his hands making sweeping gestures. He looked up in surprize when a young woman boldly challanged one of his state­ments.

THE REACTION STARTLED THE PROFESSOR. (LH)

The academician's astute observations were almost always beyond the full appreciation of his students. His keen scholarship placed him at the head of his profession, but also made his ideas a bit too com­plex. He was unaccustomed to such a delving retort as the one with which a student now confronted him.

THE PROFESSOR WAS STARTLED BY THE REACTION.
The muddy river rose higher by the hour as the rain poured down. The flood destroyed many buildings near the riverbank and completely covered some trees. The only thing that stood between the raging waters and the city was a sandbag dam, constructed earlier in the week. Finally, the rain stopped.

FORETHOUGHT AVERTED THE CRISIS. (LL)

The recession was having serious consequences in the small town. Several important businesses were on the verge of bankruptcy and the banks were unable to make adequate loans to save them. The town's entire economy might have collapsed were it not for the emergency relief fund, which the mayor had campaigned so hard to establish.

THE CRISIS WAS AVERTED BY FORETHOUGHT.

The men sitting around the table in the smoke-filled room were called to order for the vote. For the tenth time that week, no agreement was reached and the men glared across the table at one another. The chairman had wanted everyone to have an equal say in the decision, but the representatives merely argued and quibbled endlessly. In a resigned voice, the chairman announced his decision.

NECESSITY SUPERCEDED DEMOCRACY. (LL)

The Western ideal of government is an outgrowth of the Greek concept of republics. At that time, legislation was approved by direct consent of the populace. The only deviation from this process occurred in times of serious threat to the community, when a temporary tyranny was appointed.

DEMOCRACY WAS SUPERCEDED BY NECESSITY.
The tearful king made an impassioned speech, trying to justify his actions. But, at the end, and to the surprise of his ministers, he lifted the crown from his head and placed it gently on the head of his brother. The next day, in his private plane, he flew to his country estate and never returned.

AN EMBEZZLEMENT CAUSED THE ABDICATION. (LL)

Much to the surprise of his ministers, the king had relinquished his power and left the country. Recent developments had, of course, necessitated a drastic change, but few had anticipated that the king would give up his throne.

THE ABDICATION WAS CAUSED BY AN EMBEZZLEMENT.

The president knew well that he had lost the trust of his people. Outside his window he could see the angry protesters marching up and down the street carrying placards proclaiming him a liar. He turned from the window and looked sadly at the letter of resignation of his most trusted aide, who had said he could no longer work effectively under a president who would lie to a grand jury.

THE INVESTIGATION ERODED CONFIDENCE. (LL)

Trust in a nation's leadership is an important cohesive force. When this belief in the righteousness of the nation's politicians was called into question by the Watergate scandal, the feeling of deception was sufficiently generalized to cast doubt upon the motives of figures in all levels of government.

CONFIDENCE WAS ERODED BY THE INVESTIGATION.
The count stared with contempt at the soldier walking down the street. The urge to attack him from behind and kill him without warning was so strong that the nobleman found his hands shaking violently. However, despite his anger and disgust, he replaced his half-drawn sword.

**HIS HONOR CONTROLLED HIS HATRED.** (LL)

The nobleman was deeply distainful of his enemies and would have resorted to almost any means to eliminate them. This attitude stemmed from a long-standing feud and his contempt was deep and personal. However, even though it was certain to succeed, he could not agree to such a devious scheme.

**HIS HATRED WAS CONTROLLED BY HIS HONOR.**

The kings of Persia were famous as strong leaders and brave soldiers. Now Xerxes, the last of his line, was quietly layed in the stone coffin. All around in the torch-lit tomb were other magnificent coffins of his fathers. The knife of Xerxes' most trusted minister had pierced the heart of the last of the great Persian rulers.

**A BETRAYAL ENDED THE DYNASTY.** (LL)

The kingdom of Persia was ruled by one line of monarchs for almost eight centuries. The monarchs were known for their just, but absolute control. The last king was murdered after the Roman conquest.

**THE DYNASTY WAS ENDED BY A BETRAYAL.**
REFERENCES

Begg, I. Recognition memory for sentence meaning and wording. 

Begg, I., & Paivio, A. Concreteness and imagery in sentence meaning. 


Fillenbaum, S. Memory for gist: Some relevant variables. 


James, C. T. Theme and imagery in the recall of active and passive sentences. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 205-211.


Morris, P. E., & Reid, R. L. Imagery and the recall of adjectives


Pezdek, K., & Royer, J. M. The role of comprehension in learning


Tannenbaum, P. H., & Williams, F. Generation of active and passive sentences as a function of subject or object focus. *Journal of Verbal Learning and Verbal Behavior*, 1968, 7, 246-250.

VITA

PATRICK JAMES CARROLL