SOME PSYCHOLOGICAL PARAMETERS
OF VERBAL ENCODING

A thesis presented to the Department of
Psychology and Sociology
University of Canterbury

In fulfillment of the requirements
for the Degree of Doctor of
Philosophy

by

Jessie Elisabeth Wells
January, 1970
ACKNOWLEDGEMENTS

The author wishes to thank Professor R.A.M. Gregson for the supervision of this thesis. Thanks are also due to Mr. D.O. Watson and members of the technical staff for their assistance, to Miss J. Roberts for typing, and to those friends who helped with suggestions and proof-reading, in particular Mr. P.M. Russell.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>ABSTRACT</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CHAPTER 1</th>
<th>Theoretical Discussion</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Introduction</td>
</tr>
<tr>
<td>1.2</td>
<td>Definitions</td>
</tr>
<tr>
<td>1.2.1</td>
<td>Encoding</td>
</tr>
<tr>
<td>1.2.2</td>
<td>Encoding, Storage, and Retrieval</td>
</tr>
<tr>
<td>1.3</td>
<td>Labelling and Labels</td>
</tr>
<tr>
<td>1.4</td>
<td>Meaning</td>
</tr>
<tr>
<td>1.4.1</td>
<td>Reference</td>
</tr>
<tr>
<td>1.4.2</td>
<td>Imagery</td>
</tr>
<tr>
<td>1.4.3</td>
<td>Associative Processes</td>
</tr>
<tr>
<td>1.4.4</td>
<td>Emotive Processes</td>
</tr>
<tr>
<td>1.4.5</td>
<td>Multiple Aspects of Verbal Encoding</td>
</tr>
<tr>
<td>1.5</td>
<td>Words - written or spoken</td>
</tr>
<tr>
<td>2</td>
<td>Models for Encoding</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CHAPTER 2</th>
<th>Models for Encoding</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Introduction</td>
</tr>
<tr>
<td>2.2</td>
<td>Models for Memory</td>
</tr>
<tr>
<td>2.2.1</td>
<td>Sensory Registers</td>
</tr>
<tr>
<td>2.2.2</td>
<td>Short-term Store</td>
</tr>
<tr>
<td>2.2.3</td>
<td>Long-term Store</td>
</tr>
<tr>
<td>2.3</td>
<td>A General Diagrammatic Model for Encoding</td>
</tr>
<tr>
<td>2.4</td>
<td>A Diagrammatic Model for Visual Input</td>
</tr>
</tbody>
</table>

Page numbers are not consistent with the content provided.
### CHAPTER 3
**Literature Review**

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Introduction</td>
<td>40</td>
</tr>
<tr>
<td>3.2</td>
<td>Paired-Associate Studies</td>
<td>42</td>
</tr>
<tr>
<td>3.3</td>
<td>Recognition Studies</td>
<td>48</td>
</tr>
<tr>
<td>3.4</td>
<td>Free Recall Studies</td>
<td>50</td>
</tr>
<tr>
<td>3.5</td>
<td>Explanations</td>
<td>53</td>
</tr>
<tr>
<td>3.6</td>
<td>Time Parameters in Encoding</td>
<td>59</td>
</tr>
<tr>
<td>3.7</td>
<td>Summary</td>
<td>62</td>
</tr>
</tbody>
</table>

### CHAPTER 4
**Experiment 1**

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>General Considerations</td>
<td>63</td>
</tr>
<tr>
<td>4.1.1</td>
<td>Introduction</td>
<td>63</td>
</tr>
<tr>
<td>4.1.2</td>
<td>Experimental Variables</td>
<td>69</td>
</tr>
<tr>
<td>4.1.3</td>
<td>Predictions</td>
<td>70</td>
</tr>
<tr>
<td>4.2</td>
<td>Methodological Considerations</td>
<td>72</td>
</tr>
<tr>
<td>4.2.1</td>
<td>Stimuli</td>
<td>72</td>
</tr>
<tr>
<td>4.2.2</td>
<td>Response Sheets</td>
<td>73</td>
</tr>
<tr>
<td>4.2.3</td>
<td>Apparatus</td>
<td>73</td>
</tr>
<tr>
<td>4.2.4</td>
<td>Subjects</td>
<td>74</td>
</tr>
<tr>
<td>4.2.5</td>
<td>Questionnaire</td>
<td>74</td>
</tr>
<tr>
<td>4.3</td>
<td>Pilot Studies</td>
<td>74</td>
</tr>
<tr>
<td>4.3.1</td>
<td>Pilot Study 1</td>
<td>74</td>
</tr>
<tr>
<td>4.3.2</td>
<td>Pilot Study 2</td>
<td>76</td>
</tr>
<tr>
<td>4.3.3</td>
<td>Pilot Study 3</td>
<td>78</td>
</tr>
<tr>
<td>4.3.4</td>
<td>Pilot Study 4</td>
<td>81</td>
</tr>
<tr>
<td>4.3.5</td>
<td>Summary</td>
<td>83</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>--------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>4.4</td>
<td>Experiment 1</td>
<td>83</td>
</tr>
<tr>
<td>4.4.1</td>
<td>Design</td>
<td>83</td>
</tr>
<tr>
<td>4.4.2</td>
<td>Control Group</td>
<td>84</td>
</tr>
<tr>
<td>4.4.3</td>
<td>Experimental Groups - Procedure</td>
<td>85</td>
</tr>
<tr>
<td>4.5</td>
<td>Results for the Memory Task</td>
<td>85</td>
</tr>
<tr>
<td>4.5.1</td>
<td>Error Measures</td>
<td>86</td>
</tr>
<tr>
<td>4.5.2</td>
<td>Incomplete Responses</td>
<td>97</td>
</tr>
<tr>
<td>4.5.3</td>
<td>Analyses of variance</td>
<td>88</td>
</tr>
<tr>
<td>4.5.4</td>
<td>Summary of Analysis of Variance Results</td>
<td>93</td>
</tr>
<tr>
<td>4.6</td>
<td>Questionnaire Results</td>
<td>98</td>
</tr>
<tr>
<td>4.6.1</td>
<td>Open Question</td>
<td>99</td>
</tr>
<tr>
<td>4.6.2</td>
<td>Expected Percentage of Successes</td>
<td>100</td>
</tr>
<tr>
<td>4.6.3</td>
<td>Effects of the Experimental Variables</td>
<td>101</td>
</tr>
<tr>
<td>4.6.4</td>
<td>Factor Analysis</td>
<td>103</td>
</tr>
<tr>
<td>4.7</td>
<td>Discussion</td>
<td>107</td>
</tr>
<tr>
<td>4.8</td>
<td>Summary</td>
<td>113</td>
</tr>
</tbody>
</table>

CHAPTER 5
Experiment 2

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>Introduction</td>
<td>114</td>
</tr>
<tr>
<td>5.2</td>
<td>Experimental variables</td>
<td>118</td>
</tr>
<tr>
<td>5.3</td>
<td>Pilot Study</td>
<td>118</td>
</tr>
<tr>
<td>5.3.1</td>
<td>Method</td>
<td>119</td>
</tr>
<tr>
<td>5.3.2</td>
<td>Results</td>
<td>120</td>
</tr>
<tr>
<td>5.3.3</td>
<td>Discussion</td>
<td>120</td>
</tr>
<tr>
<td>5.3.4</td>
<td>Conclusion</td>
<td>123</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>5.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.4.1 Design</td>
<td>124</td>
<td></td>
</tr>
<tr>
<td>5.4.2 Stimuli</td>
<td>124</td>
<td></td>
</tr>
<tr>
<td>5.4.3 Procedure</td>
<td>124</td>
<td></td>
</tr>
<tr>
<td>5.4.4 Timing</td>
<td>126</td>
<td></td>
</tr>
<tr>
<td>5.4.5 Apparatus Problems</td>
<td>126</td>
<td></td>
</tr>
<tr>
<td>5.4.6 Questionnaire</td>
<td>127</td>
<td></td>
</tr>
<tr>
<td>5.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.5.1 30 Seconds Group</td>
<td>127</td>
<td></td>
</tr>
<tr>
<td>5.2.2 0 Seconds Group</td>
<td>134</td>
<td></td>
</tr>
<tr>
<td>5.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.6.1 0 Seconds Group</td>
<td>136</td>
<td></td>
</tr>
<tr>
<td>5.6.2 Comparison of 0 and 30 Seconds Groups</td>
<td>137</td>
<td></td>
</tr>
<tr>
<td>5.6.3 Comparison with Experiment 1</td>
<td>138</td>
<td></td>
</tr>
<tr>
<td>5.6.4 Expected Percentage Correct</td>
<td>140</td>
<td></td>
</tr>
<tr>
<td>5.6.5 30 Seconds Group</td>
<td>141</td>
<td></td>
</tr>
<tr>
<td>5.6.6 Factor Analysis</td>
<td>147</td>
<td></td>
</tr>
<tr>
<td>5.6.7 A Comparison of Encoding for Verbal and</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pictorial Stimuli</td>
<td>150</td>
<td></td>
</tr>
<tr>
<td>5.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHAPTER 6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.1.1 Comparison of Experiments 1 &amp; 2</td>
<td>158</td>
<td></td>
</tr>
<tr>
<td>6.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.2.1 Comparison with other Studies and a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Theoretical Discussion</td>
<td>160</td>
<td></td>
</tr>
</tbody>
</table>
6.3 The Use of a Questionnaire 166
6.4 Other Suggested Experiments 168
6.5 Conclusions 169

APPENDIX I 171
I.1 Types of Stimuli 171
I.2 Types of Response Templates 174

APPENDIX II 176
II.1 Instructions for Experiment 1
II.2 The Questionnaire used in Experiment 1 179

APPENDIX III 185
Relationships between error measures

APPENDIX IV 190
Treatment of the compact/distributed factor

APPENDIX V 194
V.1 Instructions for Experiment 2
V.2 Questionnaires 198

APPENDIX VI 205
Labelling used in Experiment 2

BIBLIOGRAPHY 207
Some Psychological Parameters of Verbal Encoding

Abstract

After a discussion of labelling and meaning, a diagrammatic model was proposed to illustrate possible encodings of verbal and pictorial information. Two experiments using visual stimuli were carried out to investigate the relative likelihood of these encodings. An attempt was made to equate the number of attributes encoded from each type of stimuli. Questionnaires were used to provide information on encoding. Experiment 1 required retention of a short list; verbal and pictorial stimuli were remembered equally well. Verbal encoding was predominant for both types of stimuli and was related to better recognition whereas visual imagery appeared to be nonfunctional. Experiment 2 required recall of one stimulus after an interpolated task; recall was much superior for pictorial stimuli. This was thought to be because pictorial stimuli were encoded twice, verbally and visually, to a greater extent than were verbal stimuli. Both forms of encoding were positively related to recall. These results, in conjunction with those from other studies, suggest that verbal encoding is particularly useful for sequential information.
CHAPTER 1

Theoretical Discussion

1.1 Introduction

In this thesis encoding of verbal and non-verbal stimuli is compared. The emphasis throughout is on visual stimuli; stimuli in other modalities will not be dealt with in detail. The two experiments reported herein were carried out to investigate the encoding of verbal and pictorial stimuli. Previous experiments using these two types of stimuli are reviewed in Chapter 3. To place the problem in a wider theoretical context, the nature of labelling and meaning will be briefly discussed. Labelling is important because non-verbal stimuli in any sensory modality may be named and encoded in this way as well as in terms of their sensory characteristics. Discussion of labelling leads to consideration of what words are, the ubiquitous problem of meaning, and the various ways in which they may be encoded and stored. This theoretical framework suggests the types of encoding presented in Chapter 2 in the general model for encoding and the model for visual input. It also provides
the basis, in Chapter 3, for evaluating explanations put forward by previous experimenters and makes clear certain hitherto uncontrolled for differences between verbal and pictorial material.

But first of all, what is meant by the term encoding?

1.2 Definitions

1.2.1 Encoding

The word encoding has a variety of different usages in psychology. For example, it may be used as the opposite of decoding in a discussion of coding systems (Murdock, 1968). For the purposes of this thesis it is defined as follows: "encoding" is the production of an internal representation of a stimulus (Bower, 1967a). The form of this internal representation is not specified; it might be visual, auditory, tactile and so on.

Such encoding is regarded as distinct from coding but there are problems in attempting to formalize the difference. For example, Underwood & Erlebacher (1965, p.1) define coding negatively - no coding has occurred if "the representation of the unit in memory is isomorphic to the unit as presented for learning". This absence of coding corresponds to "encoding" as used here. Unfortunately, "isomorphic" does not have a precise meaning in this context and so some other definition will be attempted. The "information" contained in a stimulus
is conceptualised as a subset of the potentially perceptible attributes of that stimulus and includes holistic characteristics, such as the meaning of a word, as well as characteristics of the individual parts, such as the typography of the individual letters. With encoding the information remains the same even though the representation may change in mode or from non-verbal to verbal form or vice-versa; with coding the information is changed. This use of encoding is not intended to deny that perception is selective but to emphasise the difference between selection and transformation.

The distinction between encoding and coding is useful in spite of its formal inadequacies because it makes clear whether the problem under discussion is one of the storage of certain aspects of a stimulus or one of "chunking" as in trigram learning (e.g. Lindley, 1966; Forrester & Spear, 1967) and other studies of coding systems (e.g. Gregg, 1967).

1.2.2 Encoding, Storage, and Retrieval

Following the practice in a number of recent articles (e.g. Atkinson & Shiffrin, 1968; Norman, 1968) the terms "encoding", "storage", and "retrieval" will be used instead of the more traditional "trace" terms such as Melton's (1963b) "trace formation, trace storage, and trace utilization".

Bower (1967a) notes that many past theories about the memory trace have postulated neurological mechanisms, as is shown clearly in Comuliki's historical review (1953); no such
mechanisms will be discussed in this thesis and it seems wise to avoid the use of the word "trace" because of its historical connotations.

These three stages, encoding, storage, and retrieval are theoretically separable but the nature of all three must be inferred from data about retrieved information. Thus, while encoding is of particular interest here, storage and retrieval processes cannot be ignored in the interpretation of the experimental results.

Memory processes do not exist in isolation. Something must exist to be encoded, stored, and retrieved; there must also be some individual within whom these processes occur. In this thesis an individual is, unless otherwise specified, an adult human with the normal sensory, motor, and intellectual capabilities of the species.

1.3 Labelling and Labels

Man is continually impinged upon by a great variety of non-verbal sensory input; only two of the senses, vision and audition, are commonly used for verbal communication. Yet it is obvious that much of this sensory input is labelled either for purposes of communication or as a way of encoding it. What is this process of labelling and how does it function?

In psychology the theories of meaning and hence of labelling have tended to be single stage theories such as that of Skinner (1957), or two-stage mediation theories such
as those of Osgood (1952; Osgood, Suci & Tannenbaum, 1957), and Housefield (1961). Apart from their other strengths and deficiencies (Chomsky, 1959; Fodor, 1965; Osgood, 1966) these theories have neglected or at least under emphasised the process of abstraction involved in the acquisition of all labels except, perhaps, proper nouns.

Even a simple label like "fireplace", which has no usual metaphorical application, is applied to a number of objects which differ in several ways. Only a small subset of the attributes of such objects can be critical for the use of the label. (It must be admitted that Osgood does acknowledge the selectivity of labelling, if not the abstraction). Depending on the demand characteristics of the situation or the past experience of the individual, different attributes will be selected as important and hence different labels may be given to the one object. (Sequi & Fraisse, 1968). For example, a particular plant in bloom in a garden may be labelled as a flower or as a violet. Labels stand for concepts, not objects. An object is appropriately labelled when it is an exemplar of the concept.

These concepts must not, however, be regarded as fixed and immutable. Some are closed in that they can be exhaustively described by an intensive listing of defining properties, but far more are open classes. Language is essentially
productive rather than repetitive; so is that part of language called naming or labelling. Contrary to a theory of conditioned responses children tend to apply new labels very widely and then learn to limit their field of designation (Lenneberg, 1967, p.332). As adults, by means of metaphor, combinations of labels, new words standing for combinations of labels, or new words coined for a new experience, humans can potentially label anything of which they are aware, although such labelling may be idiosyncratic.

In English and other Indo-European languages there is a grammatical distinction between adjectives and nouns; the latter are qualified by the former. Thus, adjectives tend to be labels for levels on one attribute, for example size, colour, or pattern, whereas nouns are usually complex multi-attribute concepts. This linguistic framework has led some philosophers to concentrate on object-noun relationships when discussing the noun "referent" (Wells, 1961, p.272) and it led Skinner (1961), as usual in the opposite direction, to a peculiar notion of reference in an ideal language in which all attributes necessary to describe an object are included in the label for that object.

Under the conditions of an ideal language, the word for "house", for example, would be composed of elements referring to color (sic), style,
material, size, position, and so on... The words for
two houses alike except for color would be alike
except for the element referring to color.... (p.234)

He goes on to say:

Such a language is manifestly impossible. Even if
we could extend the size of verbal units without
limit, the shortage of dimensions would force us at
some point to introduce nonfunctional similarities
among verbal forms and thus to violate the basic
rule. (p.235),

but this statement ignores the idea that only a small
portion of the possible description of any object may be
of any importance. That one of the main advantages for
communication and memory is that words act as simple,
arbitrary symbols for complex concepts is not a notion he
even considers. But perhaps it is foolish to treat
seriously the ideas of a man who can conceive of an ideal
language in which

Abstract responses would merely be incomplete
responses. (p.235).

Thus he regards as ideal a language which would rule out
the possibility of words such as "ideal" or concepts of
"ideal languages". Has Skinner experienced world-
weariness from an excess of theory and generalization?
Nonetheless, he is at least aware of the different number of attribute levels implied by nouns and adjectives, a point which is important in studies comparing memory for words and objects.

To summarise, labelling occurs when a word representing some concept is applied to an object, attribute, or event. It involves the selection of certain attributes as critical. Some types of labels such as adjectives involve a smaller number of critical attributes than do others, such as nouns. In spite of the variety of labels which can be given to objects a large measure of agreement can be found for members of the same language group as is evidenced by studies on colour naming (e.g. Lenneberg, 1967) and studies using pictorial material (e.g. Lieberman & Culpepper, 1965; Paivio & Yarmey, 1966).

1.4 Meaning

In labelling a word is given to some non-verbal sensory input; what processes take place when the sensory input is a word? Two situations may be distinguished, one in which the listener or reader can perceive the referent and one in which he is displaced in time or space from the referent. This latter situation is taken to include the case when there is no physical referent.

In the first situation the label serves mainly to denote some unique object. The label may be part of a
printed notice such as "Keep off the grass" or part of an utterance by some speaker. In either case the label serves to draw attention to the object which can then be perceived in all aspects by the reader or listener. Admittedly, the label may convey more than merely pointing at the object - for example, the choice of label may convey information about the speaker's reaction to the object as when someone looking at a derelict old house exclaims, "A mansion!" However, the use of the label does not prevent the listener from learning more about the object than is contained in the utterance.

This contrasts with displaced speech. In displaced speech the listener or reader is informed of only the critical attributes of the referent unless he is familiar with the object being discussed. It is in such a context, where the stimulus is a word divorced in time and place from any referent, that most of the problems of meaning have been discussed.

Because of the type of stimulus material to be used in the experiments reported in this thesis this section will deal primarily with word meaning rather than the meaning of utterances or propositions. Sometimes, for the purposes of simplicity, words will be treated as if they were independent, single units. Of course, words do not ever occur in isolation; they are always embedded in
some context be it linguistic or extra-linguistic.

Several sets of terms have been used for the components of meaning (e.g. Morris, 1946, 1964) some of which are more appropriate than others for a discussion of word meaning, particularly when the words are in lists, as in many laboratory studies, rather than in utterances. Consider Figure 1 in which word meaning is divided into components which sometimes interact. Not all of the components need have an important role in the meaning of any particular word.

![Diagram of Word Meaning Components](image)

**Fig. 1 Components of Word Meaning**

*Connotation is used here in the lay sense of the word (Weinreich, 1958).

Rommetveit (1968) has improved upon this diagrammatic representation of meaning by introducing the dimension of time, thus making it clear that meaning and its components are processes occurring within some individual. Rommetveit
is concerned with change in meaning during or following one particular presentation. This seems to fit in with Vigotsky's (1939) emphasis on the change in word meaning in the life of an individual – he also regards meaning as a complex process rather than a fixed and static state. Rommetveit's diagram (1968, p.110) is as follows:

![Diagram](image)

Fig. 2 "The word as a three-component temporal pattern: I: Input (word form); R1: act of reference, choice of specific semantic correlate; R2: process of representation (sustained); A: associative process; and E: emotive process."

The time factor has the effect of limiting or enriching meaning, depending on the duration of processing. Given a very limited time per word, or a word embedded in an utterance where all the other words also require processing, the meaning will contain less associative and emotive components – there may even be only enough time for the initial
act of reference. In a laboratory situation a person may be asked to list all the associates he can think of in one minute or to write a comprehensive statement of the meaning of a word; in such cases more remote associates, more subtle emotive processes, and more precise denotations will emerge.

1.4.1 Reference

Reference is such an important part of meaning that it has often been taken to be the whole (Wells, 1961), but Quine (1953) and many other linguists, philosophers, psycholinguists, and even psychologists have recognised the need to distinguish between denotation and connotation, reference and meaning. Ryle (1951, p.25) provides a most cogent example: "No one with a respect for sense would dream of pointing to someone......and saying......'the meaning of yonder phrase is suffering from influenza'.

Rommelvoit says that an initial act of reference determines which word meaning will be generated from a homonymous or polysemous word form (1963, p.126). He does not discuss the case of guns where the full meaning depends on at least two distinct referents being evoked or the case of words which evoke two or more referents because of the ambiguity of the context. Presumably for these words his diagram should split into two or more branches at the input, giving several initial acts of reference with their attendant emotive and associative processes if time allows.
Pollack (1969) suggests that more than one referent may be encoded for ambiguous words.

For studies of encoding and memory it is important to be aware of the possible acts of reference for the word forms used. Tulving (1966), Cofer (1967), Mandler (1967), Wood (1969) and many others have shown the importance of conceptual groupings in free recall. Such groupings are, of course, dependent on the acts of reference for the words in the lists - for example, "ball" could be clustered with "dance, hop, and party" or with "bat, gloves, and wicket". To know which act of reference has taken place is also important for studies of association, affective meaning, and even such experiments as described in Chapters 4 and 5 (c.f. 4.1.1).

1.4.2 Imagery

If the referent is a physical object or sensation, particularly if it is visual, it may evoke an image. It is most unlikely that such an image will be purely generic - if it is vivid it will doubtless contain a number of details which are compatible with the word-concept but are not defining characteristics. Titchener, of course, stands secure as one of the great imagers, able to cope even with words having no physical referents. For example, "meaning is the blue-grey tip of a kind of scoop which has a bit of yellow about it (probably part of the handle) and which is
just digging into a dark mass of what appears to be plastic material." (1909, p.19). More oddly, his images for physical objects were certainly not generic; had they themselves been stimuli, other speakers of English would not have given them the label which evoked their existence in Titchener's vivid inner world. It was not certain that his image of a triangle did join at the corners while his cow image was of a pouting rectangle. Titchener's enthusiastic introspection has done as much to discredit the imaginal theory of meaning as the inability of most people to see Locke's generic triangle. That most people do not have images for all words in their vocabulary, that those who do have images disagree, and that such images are seldom functional for identification of exemplars, all argue strongly against any imaginal theory of meaning. Nonetheless, the possibility of an image of a referent or an associate being generated upon the perception of a word should not be ignored in studies of encoding.

In an extensive review Paivio (1969) shows the importance of imagery for paired-associate learning, free recall, recognition memory, serial learning, and memory span. Imagery ratings for nouns such as those given in Paivio, Vuille, & Madigan (1968) have been found to be effective predictors of learning and memory. These ratings do not distinguish between images of referents or associates but they do confirm that words can evoke images and that the likelihood of imagery
depends on the words presented as well as on the individual who perceives them.

1.4.3 Associative processes

Tests of word association have been used in psychology since Jung (1918). Noble (1952) has even given an operational definition of meaningfulness in terms of "m" - the number of associates produced in 60 seconds - and Housefield (1961) suggested that meaning should be investigated as patterns of word associations.

Some experimenters have objected to imposing lexicographical and logical relations on associative linkages in attempts to assess similarity of associative meaning, although Deese, formerly one of the more vocal objectors, has changed his mind (Deese, 1968). There is evidence for a variety of types of association (e.g. Kaworski & Shacter, 1948; Brown & Berko, 1960; Ervin, 1961) and for the influence of age and cultural factors (e.g. Reigel, 1965).

Such studies indicate that associations may be formed in many ways, by their place in a conceptual structure, by contiguity in sentences, by contiguity in the real world, by semantic linkages, or by emotive linkages. The more probable associations for a word reflect certain cultural uniformities but these are seldom so marked that in a given situation an experimenter will be able to predict accurately what associates might be used as, say, mediating responses
in paired-associate learning or some short-term memory task.

The patterns of associations evoked by a word constitute part of its meaning. This idea of meaning as at least partially determined by verbal associations has also been applied to the "meaning" of non-verbal material (Kaworski, Gramlick, & Arnott, 1944).

1.4.4 Emotive processes

It is clear that some words contain more emotive connotations than others, even those with the same referent; for example, Chink for Chinese, Red for Communist, and nigger for Negro (although the emotive meaning of the last word has undergone rapid changes for both whites and negroes since the rise of the Black Power movement). Such words imply the attitude of the speaker to the referent as well as indicating the referent. These are relatively rare in any language but almost all words have some emotive connotations.

Osgood's semantic differential has been the main instrument used to measure emotive meaning, although he and his associates (1957) originally hoped to measure all aspects of meaning other than reference with it. In the light of criticism (Weinreich, 1958; Carroll, 1959), and further thought and research, Osgood (1962) restricted the domain of the semantic differential to affective meaning. A number of factor analyses within and across cultures and
some across modalities have shown a stable three dimensional semantic space. This space has an evaluative dimension ("good/bad"), a potency dimension ("strong/weak"), and an activity dimension ("active/passive"). However, Osgood (1963) has found that the evaluative dimension is the only one to appear in all lexical domains possibly, as Rommetveit (1968) suggests, because the adjectives used in the semantic differential scales must be denotatively irrelevant.

Thus this final component of meaning, the emotive connotations, may vary in importance and characteristics but is seldom if ever absent.

1.4.5 Multiple aspects of verbal encoding

As well as studies on association and emotive meaning, some of which have been mentioned in the preceding sections, and studies directly investigating the lexical groupings of words by sorting techniques (e.g. Miller, 1969), there are a number of other psychological experiments in which the components of meaning are particularly important.

These experiments can be viewed as memory experiments shedding light on meaning or as experiments which show that an understanding of meaning is necessary for a comprehensive theory of memory. Most of them are discussed in Wickens (1968).

Wicken's main assertions are that a single word is encoded in a rich multiplicity of ways and that many aspects
of this encoding may not enter consciousness at the time of the encoding. Evidence for the latter comes from Wickens's studies in which subjects were not aware of category changes although they responded to them and could see the changes when they were later pointed out to them. Wickens also quotes classical introspection studies by people such as Titchener, Külpe, and Moore to support the notion that not all encoding is necessarily conscious. No mechanism is suggested by Wickens but presumably some form of parallel processing would be required for the rapid encoding of a number of attributes which do not all enter consciousness. A more plausible explanation might be that the words and many of their complex inter-relations are already in long term store and are accessed when a word enters short term storage rather than all being fed in at once.

It has been found that when the Peterson and Peterson (1959) technique is used interference builds up over the first four or five trials (Keppel & Underwood, 1962) and then levels off. With this method an item is presented and must then be recalled after some seconds of interpolated activity such as counting backwards in threes or rapid colour naming. If, however, a new class of items is introduced, such as a three figure number to remember after several trigrams, recall improves. (Wickens, Born, & Allen,
1963). Wickens has seen that this can be used as something of a projective technique to uncover cognitive organization. The procedure is to present several members of one class and then a member of another class; if there is an improvement in recall as compared with recall of another member of the first class by the control group, then it may be assumed that the classes used have been encoded.

So far, using this technique, the grammatical class for single words does not seem to be encoded when the words are verbs or adjectives and when acoustic similarities such as commonclass endings are controlled for (Wickens, Clark, Hill, & Wittlinger, 1968). This finding is of dubious generality, particularly with respect to words embedded in utterances. Within and between dimension shifts for each of Osgood's evaluation, potency, and activity dimensions caused significant changes in recall level (Wickens & Clark, 1968; Turvey, 1968; Turvey, Fortig, & Kravetz, 1969) and Loess (1967, 1968) and Turvey, Cremin, & Lombardo (1969) found similar results for taxonomic changes such as names of countries followed by types of birds. The type of associated sense impression has been found to have an effect although only two dimensions have so far been investigated by Beutener (Wickens, 1968). Word frequency and word length changes also have some effect. Wittlinger (Wickens, 1968) found a consistent trend but no significant differences when he
changed from auditory to visual presentation and vice-versa, suggesting that the modality of presentation may be encoded but that it may not be as important as other aspects of the stimulus. Turvey & Egan (1969) have obtained evidence that some physical characteristics are encoded—they altered the area of the projected slide or changed from vertical to horizontal presentation of trigrams.

As Wickens notes, similar results have been found for clustering in free recall—no clustering due to grammatical class (Cofer & Bruce, 1965) and a lot of clustering due to taxonomic categories (Bousefield, Cohen, & Whitmarsh, 1958). Recognition studies by Underwood (1965), Kimble (1968), and Marshal, Rouse, & Tarpy (1969) have shown false recognition of category names or associates and false recognition based on sense impression was also found in Kimble's study. Wallace (1967) showed false recognitions to occur as a result of previous paired-associate learning, thus indicating that the false recognition phenomena can be produced experimentally.

Baddeley (1966), Baddeley & Dale (1966), Dale & Gregory (1966), Dale & Baddeley (1969), and Wickens (1968) report studies which suggest that semantic similarity is important in long term memory but less so for short term memory, the reverse being true for acoustic similarity.

Thus, these studies show that the categories of referents,
associates, and affective meaning are all related to memory for words. Physical characteristics of the stimuli and the phonetic properties of the words are also important. In both the design and interpretation of experiments using words it is necessary to remember the complex and various ways in which words are stored as part of an individual's language repertoire.

1.5 Words - written or spoken.

For normal adults only visual and auditory modes are used for language perception and communication, although people who read Braille can also use the tactual mode. Communication using Braille will not be considered here but it should be apparent that that following discussion, in many ways, could apply as well to auditory and tactual language communication as to auditory and visual language communication.

Language is acquired as an auditory form of communication by all except deaf children. By the time a child is four or five years old he will have learnt all the basics of language, both grammatical and semantic, although his vocabulary and the complexity of his sentence construction will increase as he grows older (McNeil, 1966, p.99; Lenneberg, 1964, p.79). Apart from those children who are taught to read at a very young age, the child's whole experience of language has been auditory. (Even very early
readers probably learn new words and grammatical constructions from speech rather than from writing.) When the child learns to read he must come to recognise visual patterns as symbols for the auditory patterns he already knows. But in learning to read and write words he has previously spoken and heard he is, in fact, learning symbols for symbols, forming a second link in a referent chain which now goes back from a visual pattern to an auditory pattern to whatever the auditory pattern denotes.

This poses a problem with regard to reading. When a literate person reads a word does he go backwards in the chain previously described or does he leave out the auditory stage and use the visual symbols to denote their referents directly? Similarly, when a new word is read, is it pronounced vocally or subvocally, making it into an auditory symbol as well as a visual one, or is it learnt only in its visual form? Another way of stating the problem is to ask if reading involves only visual perception or if the central mechanisms of audition and language perception, and perhaps the peripheral mechanisms as well, play some important role. Four main methods have been used to investigate this problem; the first and third can provide evidence only on the effects of the peripheral speech production mechanisms.

The first method involves recording the state of those muscles involved in speech; usually electropotentials are
taken from the tongue and the larynx although gross movements may also be recorded and, indeed, were all that could be measured before the technological advances of the 1930s and later. In the tradition of work on covert language processes such as that of Curtis (1900) and Wyczoikowska (1913), the recent studies of Faaborg-Anderson & Edfelt (1958) and McGuigan, Keller, & Stanton (1964) suggest that some degree of subvocal speech occurs during silent reading. The less competent the reader the greater the amount of subvocal speech; likewise, the more unfamiliar the material in terms of the vocabulary and style, or the particular language used, the more indications of subvocal speech even in competent readers. This suggests that new words are pronounced, albeit "silently", when they are first encountered and that the peripheral speech production mechanisms are involved to some extent in reading. However, as Osgood (1952) makes clear, there is still no satisfactory demonstration of the necessity of a motor component or evidence for any more than general peripheral activity.

Studies on tachistoscopic perception of words and psuedo-words constitute the second approach. After reviewing the literature, Neisser (1967) concludes that the sequence of visual perceptual acts corresponds structurally to the articulatory sequence and even goes so far as to say (1967, p.135), "By definition, to identify" a word or a
letter is to pronounce its name in inner or outer speech, and thus to store it in verbal immediate memory. Nonetheless he does note that with rapid "reading for meaning" short-term verbal memory may perhaps be bypassed.

The third method usually involves some short-term memory task with visual presentation. A subject sees a series of, say, eight consonants and has to say them out loud as they appear, whisper them, mouth them, perceive them silently, or produce inappropriate responses such as "the" for each consonant in order to suppress subvocal rehearsal. Murray (1965a, b, c; 1966a, b; 1967a, b; 1968), in particular, has done a number of studies in this area which confirm that peripheral speech production mechanisms play an important role in the encoding of verbal material as the more the material is voiced, the better it is recalled.

The fourth method also uses short-term memory tasks. The acoustic similarity of items is varied and substitution errors are found to occur markedly more with similar items - for example, the list GMPR might be recalled as GNBR.

Conrad (1962, 1964), Conrad & Hull (1964), Conrad, Freeman, & Hull (1965), Conrad, Baddeley, & Hull (1966), Cimbalo & Laugherly (1967), Wicketgren (1965a, b, c, d; 1966a, b, c), Adams, Thorsheim, & McIntyre (1969), and Anderson (1969) have shown clearly that letters are encoded, for short-term
memory, in terms of their auditory characteristics even when they are presented visually. Conrad (1963), Baddeley (1966), Baddeley & Dale (1966), Dale & Gregory (1966), Bruce & Murdock (1968), and Kintsch & Buschke (1969) have found the same results with words. This does not imply that the visual characteristics of words and letters are not encoded at all (c.f. Posner, 1967a) but studies such as those of Posner & Keele (1967) and Posner, Boies, Eichelman, & Taylor (1969) indicate that for some tasks auditory characteristics are more important after approximately 2 seconds.

Hintzman (1967) has criticised Conrad’s studies on the grounds that they confuse auditory and articulatory similarity and he has demonstrated that where auditory and articulatory similarity are not the same, errors are due to articulatory similarity. His article has, in turn, been criticised by Wicklegren (1969b) who questioned the assumptions used to determine what features are auditory and what are articulatory. He concluded that at present it is not possible to decide whether verbal short-term memory is auditory, articulatory, or some abstract modality.

Thus, in summary, it seems very likely that visually perceived words will be transformed into some phonetic form but it is not certain how central or peripheral this encoding will be.
CHAPTER 2

Models for Encoding

2.1 Introduction

In Chapter 1 the discussion of labelling showed the selective nature of this process which is commonly used to encode non-verbal stimuli. The discussion of meaning indicated some of the complex ways in which words may be stored in long-term memory and the various forms of encoding which can occur. In this chapter models for memory are discussed briefly and then two models for encoding are presented. The first is a very general one for all sensory input; the second deals specifically with visual input and is the basis for the subsequent literature review and experiments in which the encoding of verbal and pictorial material is compared.

2.2 Models for Memory

As Murdock (1967) notes, there has recently been a considerable degree of unanimity among model builders in the field of memory. A number of different sets of terms have been used but the basic ideas are often very similar. The models which are of importance for this discussion are

Three structural components of the memory system may be distinguished: the sensory register, the short-term store, and the long-term store (Atkinson & Shiffrin, 1968). The sensory register is sometimes called a sensory buffer or sensory store; the short-term and long-term stores are referred to as primary and secondary memory by Waugh & Norman (1965) and Murdock (1967). The separation of memory into three components is used in this chapter as a convenient framework for a brief overview. The distinction between long- and short-term stores is not of great importance for the experiments reviewed in Chapter 3 or those reported in Chapters 4 and 5. The emphasis in these chapters is on the various verbal and pictorial encodings possible for different types of stimuli.

Since Melton (1963a) the terms short-term memory (STM) and long-term memory (LTM) have often been used to refer to the type of experimental task rather than to structural components of memory. STM occurs in experiments with short retention intervals or single trials; LTM occurs in experiments with long retention intervals or multiple learning trials.
2.2.1 Sensory Registers

Most of the evidence to date on sensory registers relates only to the visual sensory buffer. To quote Kahneman (1968, p. 413), "The visual response to a brief stimulus lasts much longer than did the stimulus that caused it". While Sperling (1960) was by no means the first to suggest this, he did produce a new method of demonstrating and measuring the length of a visual sensation. Using tachistoscopic presentations of matrices of nine consonants and subsequent signals to indicate partial report he found that for signal delays of less than one second the signalled partial reports were more accurate than full reports. Although the stimulus was no longer present, subjects reported that they "read off" the signalled row of the matrix. Their experience was "perceptual" when their performance must have been from "memory".

Sperling and subsequent experimenters (e.g. Mackworth, 1963) have called the process involved an "image" but it seems reasonable to accept Neisser's (1967) objection that this word already suffers from a surfeit of meanings. Neisser suggests the term "icon" and goes on to talk of "transient iconic memory" or just "iconic memory". This terminology has been accepted by some other experimenters (e.g. Newholt, Nerikle, & Bryden, 1969).
Studies of masking have further clarified this phenomenon and shown how the duration of iconic memory depends on the intensity of both the test and the masking stimuli, exposure time, the type of material, and the post-exposure field (Kahneman, 1968). Sperling's delayed signal method indicates that iconic memory can last for less than a second if followed by a bright exposure field but that with a dark post-exposure field it may last for up to five seconds (Sperling, 1960, 1963; Averbach & Sperling, 1961).

Although there are many problems to be solved in masking studies and even technical reasons to doubt the accuracy of Sperling's reported presentation times (Priest & Crowther, personal communication, 1969), there is no reason to doubt that there is some kind of transient visual store.

The nature of an auditory sensory buffer is less clear (Atkinson & Shiffrin, 1968) but Neisser (1967) presents evidence for an "echoic memory" with a duration of one to four seconds. Such a sensory store seems essential for speech and rhythm perception. Wicklegren (1969a) also provides evidence for some very short-term auditory register.

Aaronson (1967) reviews a number of studies in which iconic and echoic memory considerations have been neglected
and emphasises that perceptual accuracy has not always been ensured in short-term memory tasks. Unfortunately little is known about sensory registers in other modalities.

The stimulus information in the sensory registers is read out into short-term store by pattern recognisers and stimulus analyzers, as Bower (1967a) calls them. The amount read out from each register will depend on the processing demands and "attention" of the perceiver (Murdock, 1967). The result of the pattern recognisers and stimulus analyzers is some kind of perceptual synthesis (Neisser, 1967) based on generative programs in long-term store.

2.2.2 Short-term Store

Theorists in the verbal learning tradition such as Melton (1963a, b) and Postman (1964) have objected to dichotomising memory into long- and short-term components. However, their arguments for the parsimony of a one memory system are not always supported by comparison of their theoretical discussions with some of the elegant two state models. Also, Melton's (1963b) objection to the apparently indeterminate point at which short-term storage ends and long-term storage begins is nullified in models in which items may be in either or both long- and short-term store (Waugh & Norman, 1965). Nonetheless it is difficult in many experiments to distinguish between long- and short-
term components of memory.

It is not clear whether or not there is a short-term store for each modality. The discussion in section 1.5 on the encoding of words presented much of the evidence for a short-term verbal memory called "active verbal memory" by Neisser (1967), "auditory information store (AIS)" by Sperling (1967), and "auditory-verbal-linguistic short-term store (a-v-l STS)" by Atkinson & Shiffrin (1968).

Items may be retained in short-term verbal memory by rehearsal; rehearsal may also be the means of transfer to long-term store. Sperling (1967) and Bower (1967a) discuss in detail the function of rehearsal for the short-term verbal store. Neisser (1967, p.224) has suggested that rehearsal is basically the same sort of synthesis as takes place during perception. If this is so then a similar sort of rehearsal would presumably be possible in other short-term stores but there is little evidence as yet.

The work of Posner (1966, 1967b) and Posner & Konick (1966) indicates a short-term visual store with some rehearsal capability and a kinaesthetic short-term store without rehearsal. Cohen & Granström's (1968) results also suggest a visual store. Evidence for rehearsal comes from comparison of recall after rest and recall after some dissimilar interpolated activity; if recall is worse after an interpolated task then presumably some rehearsal took
place during the rest period. Only Williams, Beaver, Spence, & Rundell (1969) have been able to show any interference effects in kinaesthetic short-term memory even from an interpolated kinaesthetic task. Using similar methods, Gilson & Baddeley (1969) have found evidence for what may be a tactile sensory register from which information decays and a less labile tactile store which has some non-verbal rehearsal.

The studies involving imagery reviewed in Paivio (1969) and those comparing memory for pictures and words reviewed in Chapter 3, also indicate that visual stores, whether long- or short-term, may be more important than has been thought previously. Work on memory has been carried out with mainly verbal stimuli and this has lead to neglect of non-verbal encoding.

Transfer from short-term to long-term store may be assumed to take place automatically, by rehearsal, regrouping, restructuring, or coding.

2.2.3 Long-term Store

Most experiments testing for recall from long-term store have used verbal material but the ability of subjects to recognise smells, tastes, and so on indicates some long-term store for each sensory modality (Atkinson & Shiffrin, 1968). Non-sensory information such as temporal sequence may be stored (Yntema & Trask, 1963). There is also
evidence for a visual long-term store (e.g. Bahrick & Boucher, 1968; Nickerson, 1968). Unlike the sensory register from which information decays rapidly, or the short-term store from which information is eventually lost, the long-term store is relatively permanent. Indeed, it is difficult to distinguish between loss of information from this store and inability to retrieve what is stored.

2.3 A General Diagrammatic Model for Encoding

This model folds out to the right from behind page 36 and should be opened for viewing while this section is read.

The model is an attempt to summarise possible encodings and makes no claims to completeness. It is exceptionally vague about storage, not even specifying whether this is long- or short-term. Nor does it provide any means for retrieving information from storage but presumably this is permissible in a model for encoding, if not in a model for memory. Also, search and retrieval mechanisms are difficult to portray diagrammatically. The generality of the model means that treatment of the processing and encoding of any particular input is summary. For example, the various and complex storage of words discussed in Chapter 1 is referred to merely as "storage of semantic characteristics".
Primarily the model indicates that sensory input may be stored in terms of its sensory characteristics or in terms of labels. Unfortunately the clear cut choice points (the diamonds) to not permit multiple encoding; this is the major fault of the model as there is no reason for encodings to be mutually exclusive. A visual pattern may be named and yet its visual characteristics may also be encoded.

The choice point labelled "LANGUAGE" is one at which the encoder must decide whether the incoming information is verbal or not. It seems probable that this choice point does occur early in processing - most people are familiar with the experience of seeing a spine of a book or a distant poster and trying unsuccessfully for some time to read what is on it before realizing that there is merely a stylistic decoration. The converse may also happen - one may suddenly realize that the pattern actually forms a word. Such experiences provide some evidence that there is a "language/not language" decision to be made prior to perceptual synthesis but they also suggest that if no "sensible", acceptable synthesis is formed then one may return to the choice point, reverse the previous decision, and try another synthesis. No provision is made for this in the model.

It was decided to concentrate on visual input for a
comparison of the encoding of verbal and non-verbal input. There seem to be more words relating to objects or qualities perceived visually than to those perceived auditorily and there are already a number of studies comparing verbal and non-verbal visual input (see Chapter 3). Following Rohwer, Lynch, Levin, & Suzuki (1967) and Otto, Koenke, & Cooper (1968) the word "pictorial" will be used to denote any visual stimulus which is not a word form; thus it could apply equally well to an object, a slide of an object, or a line drawing. It is worth noting that while a verbal stimulus is always a symbol a pictorial stimulus may or may not be. For example, a drawing of a house is a symbol of some house, real or imaginary, but a drawing of a trapezium or the presentation of a cat involves not a symbol but an actual trapezium, a real cat. For this reason the terms "verbal" and "pictorial" seem more appropriate than Underwood's (1952) terms, "perceptual" and "symbolic". The use of "verbal" and "visual" for words and pictures which are both presented visually (Kaplan, Kaplan, & Sampson, 1968; Sampson, 1969) is confusing and will be avoided.

2.4 A Diagrammatic Model for Visual Input

So that this model may be viewed while the text is read it should be folded out from behind page 38.
VISUAL INPUT

ICONIC MEMORY

NO

LANGUAGE

YES

STORED IN TERMS OF VISUAL PHYSICAL CHARACTERISTICS

NO

NAMED

YES

STORED AS WORDS OR AS PHRASES

NO

NAMED

STORED IN TERMS OF AUDITORY PHYSICAL CHARACTERISTICS

YES

LANGUAGE

YES

ECHOIC MEMORY

STORED IN TERMS OF AUDITORY ARTICULATORY AND SEMANTIC CHARACTERISTICS

NO

NAMED

STORED IN TERMS OF SENSORY CHARACTERISTICS

AUDITORY INPUT

ECHOIC MEMORY

OTHER SENSORY INPUT - MOTOR, Olfactory, etc.

SENSORY BUFFER?
In the model solid lines represent highly probable paths and dotted lines represent less probable ones. Perhaps the branching technique used to indicate multiple encoding could be replaced by a series of extra choice points but it seems impossible to do this and still keep the various encoding alternatives at the same level within the diagram. If branching can be replaced by extra choice points it certainly leads to a very much more complicated diagram.

The "language" choice point may not be used when there is correct prior knowledge of the type of input, as has been assumed for this model, but the storage of such knowledge may best be represented by the prior setting of a choice point. Hence in the model there is only one line leading out of each of the language choice points. The problem of return to the "language" choice point, discussed for the general model, cannot arise with correct prior knowledge of the type of input.

The storage in the model is divided into long- and short-term components but in most experiments comparing verbal and pictorial material both components are involved and so this distinction will not be consistently maintained. The use of the word "auditory" or the compound "auditory/articulatory" does not imply any decision as to the nature of short-term verbal memory - it is impracticable to always
list Wicklegren's (1969b) three alternatives. (1.5).

It is important to note that the diagram implies that:

i) pictorial stimuli may be stored in terms of their visual characteristics, regardless of whether or not they have been labelled, and

ii) verbal stimuli may likewise be stored in terms of their graphical characteristics, regardless of whether or not the stimulus words are transformed into some phonetic form in short-term memory.

To illustrate that it might be possible for a word to be perceived visually and stored in terms of its semantic characteristics without being subvocalized, a dotted line extends from the auditory transformation choice point to the long term storage of semantic characteristics.

The three possible forms of encoding for each type of stimulus are summarised in the following table.

<table>
<thead>
<tr>
<th>Pictorial Input</th>
<th>Verbal Input</th>
</tr>
</thead>
<tbody>
<tr>
<td>i) storage of the visual characteristics of the stimulus</td>
<td>i) storage of the visual characteristics of the word</td>
</tr>
<tr>
<td>ii) phonetic/semantic storage of the label</td>
<td>ii) phonetic/semantic storage of the label</td>
</tr>
<tr>
<td>iii) visual storage of the label</td>
<td>iii) visual storage of the referent or an associate</td>
</tr>
</tbody>
</table>
Associative images or synaesthesias in any modality other than visual are not portrayed in this model. None of the experiments reviewed in Chapter 3 suggest that these might be important.
CHAPTER 3

Literature Review

3.1 Introduction

In Chapter 2 the possible encodings for verbal and pictorial stimuli were shown in a diagrammatic model. In this chapter studies comparing memory for verbal and pictorial material are reviewed. Concept learning experiments (e.g. Runquist & Butt, 1961) are not included and nor are studies using labelled versus unlabelled pictorial material (e.g. Kurtz & Hovland, 1953). All the studies listed in the following summary found evidence to indicate superior recall or recognition for pictorial rather than verbal material but those asterisked had one experimental condition or group of subjects for which this superiority was not shown. Thus, overall, pictorial material appears to be remembered better than verbal material. These studies will be discussed in more detail and explanations will be evaluated.
Studies comparing Memory for Verbal and Pictorial Stimuli

Paired-associate Learning

Winer & Lambert (1959)
Epstein, Rock, & Suckerman (1960)
Paivio & Yarmey (1966)
Ronner, Lynch, Levin, & Suzuki (1967)
Otto, Koenke, & Cooper (1968)
Deno (1968)
Dominowski & Gadlin (1968)

* Dilley & Paivio (1968)

Recognition Studies

Jenkins, Neale, & Deno (1967)
Shepard (1967)

* Dallett & Wilcox (1968)
Fozard & Lapine (1968)
Fozard (1968)
Coraini, Jacobus, & Leonard (1969)

Free or Ordered Recall

Lieberman & Culpepper (1965)

* Ducharme & Fraisse (1965)
Scott (1967)
Paivio, Rogers, & Smythe (1968)
Kaplan, Kaplan, & Sampson (1968)
Sampson (1969)

* Paivio & Csapo (1969)
3.2 Paired-Associate Studies

The paired-associate studies fall into three groups; those varying only the stimuli, those covarying stimuli and responses, and those varying stimuli and responses independently.

Studies in which verbal and pictorial items serve as stimuli for nonsense syllables or numbers were carried out initially because they were relevant to the word-word versus picture-word controversy in foreign language vocabulary learning. Some of the early studies are reviewed in Winer & Lambert (1959) but they were methodologically inadequate and do not merit further discussion. Winer & Lambert's (1959) study was the first important one in this area.

They had subjects learn nonsense syllables as responses to words or pictures. Significantly fewer errors were made in the acquisition of the object-syllable list than with the word-syllable list. In a second experiment they measured Noble's m, the number of written associates produced in 60 seconds, and found that on this criterion of meaningfulness, there was no difference between the word list and the object list. However, using Osgood's semantic differential they found the average inter-stimulus semantic distance to be greater for the object list and so concluded that their results were due to the lower intralist
similarity of the object list. They did not suggest why objects should be less similar than their labels.

Dominowski & Gadlin (1968) used three types of stimuli—pictures, object names, and category names. Stimulus materials did not differ in meaningfulness (m) or the proportion of unique associations but significantly more noun associations were given to the category names than to the pictures or the object names. The object names had shorter latencies for image arousal than did the category names. Pictures were learnt most quickly, then object names, then category names.

The object name/category name difference was replicated in a second experiment but presentation of the appropriate picture together with the word on the first study trial did not facilitate the acquisition of either list. This was interpreted, in conjunction with subjects' reports, as evidence against an imaginal mediation hypothesis. The order of paired-associate learning—pictures, object names, and then category names—was thought to be due to the greater differentiation among more concrete stimuli. As with Winer & Lambert (1959), no reason was given to account for such differences in differentiation, and hence intralist interference.

Using a Peterson & Peterson (1959) technique, but with recall required for both members of the paired-associates,
no differences in forgetting due to stimulus type were found for the conditional recall of number responses (i.e. recall of a response given that the stimulus item was correctly recalled). This was thought to support the differentiation hypothesis by implying that the imagery level of the stimulus serves only to differentiate between stimuli and does not aid recall of one item. The presentation time per pair was only 0.3 seconds and the picture-number pairs were perceived more poorly than the word-number pairs, as indicated by immediate recall. Nonetheless, at 12 seconds words and pictures were recalled equally well. This suggests that those pictures which were perceived correctly were remembered better than the words which were perceived correctly, contrary to Dominowski & Gadlin's interpretation. The drop in recall for pictorial material at the 3 seconds retention interval is unaccounted for.

Deno (1968) varied the conceptual similarity of his stimulus lists as well as having verbal or pictorial presentation of stimuli in his language learning experiment which required the acquisition of Japanese words as responses. None of his Ss knew any Oriental language. For all lists, and on all measures of learning, pictorial stimuli were superior to verbal stimuli although the difference was not significant for the dissimilar lists. Deno suspected that these results were due to differences in meaning between
pictures and words even though the pictures were readily identified by the appropriate labels. As evidence, he quoted a study by Deno, Johnson, & Jenkins (1968) who found some difference in patterns of association for verbal and pictorial stimuli.

The first study to cover both stimulus and response types was one by Epstein, Rock, & Zuckerman (1960) in their monograph on meaning and familiarity in paired-associate learning. As in all of these studies in which response items are presented verbally or pictorially, the subjects' responses were always verbal. They compared object-object, concrete noun-concrete noun, abstract noun-abstract noun, and verb-verb lists. Object pairs were learnt more easily than their referents, the concrete noun pairs, which were learnt more easily than the verb or abstract noun pairs, the last two being of equal difficulty. The explanation offered was that of Kohler's organization hypothesis, the idea being that somehow it was easier to build a unified concept out of concrete noun pairs than out of abstract noun pairs. Object pairs and verb pairs were just included as possibly interesting conditions.

The effect of verbal connectives on paired-associate learning in children has been investigated in studies using pictorial material (Reese, 1965; Rohwer, Lynch, Suzuki, & Levin, 1967; Milgram, 1967) but only Rohwer, Lynch, Levin,
& Susuki (1967) have compared verbal and pictorial material in this context. Using 3rd and 6th grade children and object or name pairs they found a complicated effect from connectives but more efficient learning with pictorial material for both age groups.

Otto, Koenke, & Cooper (1968) found less errors with pictorial material for both good and poor readers in the 2nd and 5th grades.

The only studies to vary stimulus and response type independently are those of Paivio & Yarmey (1966) and Dilley & Paivio (1968). Paivio & Yarmey took particular care with selection of material using only those pictures which twenty five judges found easy to name and for which there was little variation in the label given to the picture. They found a significant difference between stimulus types, pictures being easier to learn than words, but no response type effect. Picture-noun pairs were easier to learn than picture-picture pairs possibly, they suggested, because more time is required to decode a picture memory image into a verbal response form than to recall a noun. The stimulus effect gives further support to the "conceptual peg" hypothesis which has been found useful as an explanatory concept in studies of paired-associate learning in which meaningfulness, concreteness, and rated imagery have been varied (Paivio, 1969). The hypothesis states that a
concrete stimulus member may function as a "conceptual peg" in that, when it is presented alone, the appropriate response term may be retrieved by means of a mediating image. The image mediation hypothesis has been criticised by Dominowski & Gadlin (1968). In this framework, pictures can be regarded as one end of a continuum of concrete imagery with abstract words at the other extreme.

Dilley & Paivio (1968) used nursery school, kindergarten, and 1st grade children and so they had to present words auditorily as not all the children could read. In spite of this inevitable confounding of mode of presentation and type of material they obtained results consistent with those of Paivio & Yarmey. The only difference, that there was significantly poorer recall for picture response items than for word response items, was predicted on the grounds that children would have greater difficulty than adults in decoding a memory image of an object into the appropriate response, even though their ability to name the pictures had been tested previously.

These studies generally show that paired-associate learning is superior for pictorial material although the studies of Paivio & Yarmey (1966) and Dilley & Paivio (1968) suggest that those studies which vary stimulus and response type together show pictorial superiority merely because of the stimulus type. The superiority of pictorial material
appears to hold for adults and children, objects, photographs and coloured or uncoloured line drawings, and for immediate recall trials and trials to criterion.

3.3 Recognition Studies

These are a heterogeneous group of studies on list recognition, continuous recognition, and recency.

Jenkins, Neale, & Deno (1967) presented 17 items which Ss then had to recognise from a series of 44 items. Presentation and recognition items were verbal or pictorial, giving four conditions - PP (pictures-pictures), PW (pictures-words), WP, and WW. On the number of correct recognitions, PP was superior to WW, there was no difference between WW and PW, and both were superior to WP. Confidence ratings followed the same pattern.

Dallett & Wilcox's (1968) study was different from the others discussed in this chapter in that the pictures were more complex and had corresponding descriptions 40 - 100 words long instead of one word labels. One group of Ss heard descriptions of complex pictures and were then given a recognition test with pictures which did or did not correspond to the descriptions. This order was reversed for a second group who saw a picture and then were tested with a changed or unchanged description. There were no differences between the groups for presentation times of 10 and 20 seconds but the "picture-description" group did
poorly with only 5 seconds presentation time. Presumably these Ss did not have time to encode all that was in the picture and so missed some of the features which were salient for recognising if the following description was apt or not. In a second experiment, mixed lists of pictures and descriptions were presented and tested after 1 minute, 2 days, or 1 week. Recall was significantly better for pictures than for descriptions but recognition scores, obtained after recall (c.f. Belbin, 1950), were complicated and difficult to interpret. Dallett & Wilcox suggested that the pictures were learned better but remembered no better as the slope for recall declined at the same rate over time for both pictures and descriptions but they did not define what they meant by learning and remembering nor did they give distinguishing criteria.

Shepard (1967) had Ss look through a series of six hundred items and then tested recognition for "old" stimuli in a series of test pairs. Both the original and test series were the same type of stimuli. Pictures were recognised better than words, and words better than sentences. There was no correspondence between the three types of stimuli - for example, the words did not label the pictures.

In two recency experiments Pozard & Lapine (1968) and Pozard (1968) found better discrimination and better
recognition for pictures than for words.

Using 5 year old pre-school children Corsini, Jacobus, & Leonard (1969) compared recognition memory for words and pictures. For each item in the word or picture series the children had to indicate if it was new or a repeat of a previous item. The very significantly lower error rate for pictures confirmed that children remember pictures better than words. Ducharme & Fraisse (1965) and Dilley & Paivio (1968) found children to do no better or even worse with pictorial material for which a verbal response was required; they interpreted this as evidence that children had difficulty decoding a memory image rather than that the pictorial material was retained less well. The study by Corsini et al. supports their interpretation.

Thus, recognition memory appears to be better for pictures than for words over a variety of experimental tasks.

3.4 Free Recall Studies

Apart from two studies late last century, mentioned by Paivio (1969), the free recall studies comparing verbal and pictorial stimuli began in 1965 with the experiments of Lieberman & Culpepper and Ducharme & Fraisse.

Lieberman & Culpepper (1965) presented a list of words or objects once and then asked for free recall. With total time constant but with time for each item not controlled,
and with both total time and item time controlled, they found superior free recall for objects, or slides of objects. As in all of these free recall studies, the words named the objects.

Ducharme & Fraisse (1965) used lists of words, pictures, and words plus pictures and found a slight trend for children to find the words easier and for adults to find the pictures easier. They considered that this was because a written verbal response was required so that children, who verbalized less spontaneously than adults, were helped when an appropriate response was provided. However, the general lack of significance in this study makes it difficult to come to any definite conclusions about it apart from the clear cut recency effects for pictures and primacy effects for words which occurred at all age levels.

Scott (1967) found significantly higher recall for objects than for words and also significantly more clustering in the objects list. However, Paivio, Rogers, & Smythe (1968) obtained higher recall for pictures without more clustering so that the superior recall for pictorial material cannot be due entirely to greater inter-list organization.

Paivio, Rogers, & Smythe (1968) presented nouns, black and white line drawings, or coloured drawings for four free recall trials. As Ducharme & Fraisse found, there were
serial position differences between the types of material, pictures being superior for terminal input items on Trial 1 and for both terminal and early items on Trial 2. Paivio et al. suggested three reasons for their results: i) pictures are less susceptible to serial interference; ii) they are better stored in long-term memory and are thus more readily recalled from short-term memory; or iii) pictures are stored both visually and verbally. There was no significant difference in recall scores between the coloured and the uncoloured drawings which was evidence against Bousefield, Paterson, & Whitmarsh's (1957) hypothesis of the compounding of signs, each one being conditioned to a response so that the more signs in a stimulus, the better it will be recalled. Bousefield et al. had found recall for coloured drawings plus names to be higher than that for uncoloured drawings plus names which was, in turn, higher than that for names alone. As they did not analyse for the amount of clustering their results could have been due to the organization of the lists which could be divided into five categories on the basis of meaning (birds, fruits, etc.) or colour (white, yellow, etc.). The possibility of grouping by colour would have been obvious only with the coloured drawings.

Kaplan, Kaplan, & Sampson (1968) and Sampson (1969) used a rather different experimental procedure. Each S
saw a list which was half words and half simple line drawings with two slides between each stimulus item. These intervening slides contained coloured dots which Ss had to name. GSRs were measured during presentation and during recall which consisted of reproduction of the original stimuli. In both studies recall was significantly higher for the pictorial material immediately after presentation and one day later, and under incidental or intentional learning conditions. The role of arousal was difficult to interpret as the results of the first study were not replicated in the second.

In summary, when adult subjects are used, free recall is superior for pictorial material even when a verbal response is required.

3.5 Explanations

What can account for the number of studies showing that pictorial material is easier to remember than verbal material, be the task paired-associate learning, free recall, or recognition? Winer & Lambert (1959) and Dominowski & Gadlin (1968) found no differences in meaningfulness, as measured by Noble’s w, but Deno (1968) thought that there might be differences in meaning for verbal and pictorial material in terms of the associations evoked. Kaworski, Gramlick, & Arnott (1944) found only slight differences in patterns of association and these seem inadequate to account
for those studies with clear evidence of better retention for pictorial material. Also, while Dominowski & Cadlin (1968) did find more noun association responses to category words they found no difference between the proportion of noun associates given to objects and to their names.

Jenkins, Heale, & Deno (1967), Paivio, Rogers, & Smythe (1969), Kaplan, Kaplan, & Sampson (1969), Sampson (1969), and Corzini, Jacobus, & Leonard (1969) all suggest that words are encoded only verbally while pictures are encoded both visually and verbally. Jenkins et al. point out that there is little social reinforcement given for remembering the visual characteristics of words such as the kind of printing type used or whether the word was in upper or lower case. In section 1.5, evidence was presented for an auditory, articulatory, or at least phonetic encoding of visually presented words. The genetic reason given for this was that language is learnt first as an auditory, speech communication system and then written words are learnt as visual symbols or substitutes for spoken language. Thus the notion that pictures are encoded both visually and verbally and words only verbally seems a very reasonable one. It is, in fact, implied in the diagrammatic model for visual input (Fig. 4, 2.4) where there are two main forms of encoding for pictorial material (represented by solid lines), visual and verbal, but only one main form for words, that is,
verbal. Storage of the actual visual characteristics of words or storage of visual images evoked by the words is considered less likely and so is represented by dotted lines. Of course the probability of these forms of encoding will depend on the task, for doubtless subjects can be trained to remember visual characteristics of words if required to do so. It will also depend on the particular words used for, as Paivio (1969) has repeatedly shown, words differ widely in the amount of imagery they evoke.

However, there is one other major difference between verbal and pictorial material or, more specifically, between an object and the name given to that object. The chapter on labelling and meaning should have made it quite clear that names are concepts and that any object, or picture of one, can be only a member of the class of exemplars of a concept. As such, it will have attributes which are not included in the defining attributes of the concept. For example, any cup will be a certain size, thickness, and colour so that a list of adjectives can be placed in front of the word "cup" to more fully describe that particular cup. The appropriateness of the label "cup" is not in doubt - all people over the age of five who speak fluent English may give that particular object the label "cup" - but labelling is not giving a full description; there are cups and cups. Therefore a larger
number of attributes is carried by a list of objects or pictures than by a list of names. Lieberman & Culpepper (1965), for example, found that subjects who had seen pictures tended to report adjectives as well as the nouns required for free recall, giving evidence that at least some subjects had encoded more than just names.

This argument can be extended to comparison of objects, pictures, and line drawings. Using normal and retarded children, Iscoe & Semler (1964) found paired-associate learning for objects to be superior to that for coloured photographs of the objects. They attributed their results to cue reduction in the photographs.

Unfortunately, it is difficult if not impossible to determine in advance exactly how many attributes any one person will encode from a particular object or picture. The most important factor is probably the context - if all cups are the same size, size may not be encoded - but the task doubtless also influences encoding - for example, Paivio, Rogers, & Smythe (1968) did not require recall of colour and so Ss may have regarded colour as relatively unimportant. Bower (1967a, b) has suggested in the hierarchical version of his multicomponent model that attributes may be stored in order of importance with the least important being forgotten first.

Neisser (1967) notes that some aspects of stimuli
seem to create storage rather than to fill it up and Norman's (1968, 1969) discussions of retrieval suggest that the more tags there are on an item, the more readily it will be found. Farrimond (1968) regards the extra integrated cues in a picture as aids to retention. Thus, probably one of the main reasons why pictures are remembered more easily than words is because they are more richly encoded and are therefore more readily retrieved.

Jenkins, Neale, & Dargo (1967) put forward an argument similar to this, saying that pictures were more "distinctive" than words. Murdock (1960) has noted that the concept of "distinctiveness" is an intuitively satisfactory one as an explanation for a variety of phenomena but that it is seldom clearly defined. His model for distinctiveness along one dimension cannot be applied here but some measure of the number of attributes involved and their relative importance would seem to be what Jenkins et al. are talking about. They also point out that the extra encoding for pictures may be linguistic or nonlinguistic.

The compounding of signs idea of Bousefield et al. (1957) is similar to "distinctiveness" although the vocabulary of classical conditioning is not used in the latter.

Scott (1967) also talked of the greater abundance of cues provided by pictorial stimuli but he did not discuss why this was so.
The greater semantic distance found between objects by Wimer & Lambert (1959) and Dominowski & Gadlin's (1968) differentiation hypothesis both fit within the explanation of pictorial superiority as a result of the encoding of more attributes from pictorial stimuli. However, this explanation predicts better recall for single stimuli as well as for lists which is contrary to Dominowski & Gadlin's view. It is hard to see why any factor which aids recall of a list should not also have a facilitative effect on single items as well. The separation in time between single items and items in a list is quantitative, not qualitative.

Even the "conceptual peg" hypothesis may perhaps be partly explainable in terms of the number of attributes which can be encoded from different types of stimuli.

In a way, Dallett & Wilcox (1968) did acknowledge the extensive amount of information available in pictures by comparing recall and recognition for pictures with that for lengthy descriptions but their material was so complex that it is difficult to tell how much of the information in the pictures was contained in the descriptions.

There is still a need for a study using simple material in which the number of attributes carried by words and by pictures is approximately the same. This could perhaps be done by using stimuli which are more or
less one dimensional such as colours differing only in wavelength but if more than four or five colours must be discriminated the verbal labels become clumsy. The same problem arises with stimuli varying only in size or brightness. Therefore, carefully selected multi-attribute stimuli appear to be the most suitable. For example, Ss could be instructed to recall size, colour, and shape and be presented with a large red circle or the words, "large red circle". With such stimuli the hypothesis that pictures are stored both verbally and visually but that words are stored only verbally can be investigated with less confounding of the number of attributes stored and the number of forms of storage of the same attributes.

3.6 Time Parameters in Encoding

In all of the studies discussed so far there was ample time for the encoding of several aspects of the stimuli, the naming of pictures, or image arousal from words. Presentation times of 4 or 5 seconds per pair were used in most of the paired-associate studies and inter-trial intervals of 2 to 3 seconds were used in most of the free recall experiments. What results would be obtained at very much faster presentation rates? Paivio & Caspo (1969) investigated this in two rather elegant experiments in which they manipulated the availability of non-verbal imagery and verbal mechanisms as memory codes.
They noted that pictures, if familiar and unambiguously labelled, readily evoke both concrete imagery and verbal encoding but that verbal encoding is less available because an extra transformation is involved; it is even less readily available if the pictures are unfamiliar or ambiguous (c.f. Oldfield & Wingfield, 1965). Fraisse (1964) has shown that words can be read faster than objects can be named and Paivio (1966) has shown that latency of image arousal, as measured by a key press reaction time, is faster for concrete than for abstract nouns. Comparison of the reaction times quoted in the different experiments indicates that image arousal to words is slower than verbal encoding of words or familiar pictures. Paivio & Csapo assumed that it takes no longer to perceive a familiar picture than to read a word implicitly; Fraisse (1963) has shown that motor reaction times and recognition thresholds may actually be a little lower for pictures. Thus, Paivio & Csapo concluded that summed availability of both visual and verbal encoding is greatest for pictures, intermediate for concrete nouns, and lowest for abstract nouns, although this ranking can be altered by increasing the rate of stimulus presentation.

They also noted that it is reasonable to assume that, as visual imagery is related to visual perception, it will be primarily a parallel processing system whereas the verbal symbolic system can be assumed to be specialized
for serial processing. Hence, one would expect performance in nonsequential memory to depend on the availability of both visual and verbal encoding but performance in sequential memory to depend on the availability of verbal encoding alone.

In their two experiments availability of imagery was varied by the use of abstract words, concrete words, and line drawings of the objects labelled by the concrete words. The availability of verbal encoding in the case of pictures was varied by presenting the stimuli at rates presumed to be above and below labelling threshold. The fast rate was 5.3 items per second and the slow rate of 2 items per second was four to six times faster than the rates used in the studies already discussed. Immediate memory span and serial learning were the sequential tasks and free recall and recognition were the nonsequential tasks. As had been predicted a) memory for pictures was significantly inferior to words only in the sequential memory tasks, and then only at the fast rate; b) both pictures and concrete words exceeded the abstract words in serial learning at the slow rate; and c) pictures were significantly superior to the abstract words at the slow rate in both nonsequential tasks, with concrete words intermediate in each case.

This study shows that both rate of presentation and the type of task affect encoding.
3.7 Summary

There are two main explanations for the superior recall and recognition of pictorial material:

i) given sufficient time pictures are encoded visually and verbally but words are encoded only verbally unless they evoke some imagery

ii) because pictures are more concrete than words they carry more attributes; thus they are encoded more richly and can be retrieved more readily.

These explanations are not mutually exclusive but so far the evidence for i) has been confounded by ii). There is a need for an experiment which attempts to control for the number of attributes encoded from each type of stimulus. Paivio & Csapo's (1969) study shows that encoding also depends on the nature of the task and the rate of presentation.
CHAPTER 4

Experiment 1

4.1 General Considerations

4.1.1 Introduction

The aim of this experiment was to compare memory for verbal and pictorial material when approximately the same number of attributes are encoded from each type of stimulus and also to explore the types of encoding suggested by the diagrammatic model for visual input (Fig. 4, 2.4).

As discussed in section 3.5, stimuli varying on only one attribute are unsuitable because of the small number of one word labels per attribute. The most appropriate sort of pictorial stimuli seem to be the geometric ones used in many concept learning experiments (e.g. Bourne & Restle, 1959; Haygood & Bourne, 1965) and by Lawrence & LaBerge (1956), Harris & Haber (1963), and Haber (1964 a,b). Unlike photographs of actual objects which have a large number of features which can be described if sufficient time is given (e.g. a candle which hasn't ever been lit, about 10" high, made of plain, white wax....) or a drawing in which the characteristics of the drawing may be encoded as well as
characteristics of the object drawn (e.g. a well rounded banana, drawn with clear black contours and no shading...), coloured geometric figures are usually described more simply. For example, Harris & Raber (1963) used stimuli such as "two red stars". Admittedly, background colour, illumination, and the relative positions of the stars could also be encoded but if these are always the same, presentation time is limited, and Ss are specifically instructed to recall only number, colour, and form, it does seem likely that only these three attributes will be encoded. Experimental studies of spontaneous descriptions of different types of material would be useful, although it is clear that results would be very dependent on the instructions given and might not reflect the richness of encoding, particularly if some features of the stimuli were encoded nonverbally. In spite of the difficulties involved, it seems reasonable to expect Ss to encode roughly the same number of attributes from, say, a picture of a large red circle and the words "large red circle" whereas one would expect widely different numbers of attributes to be encoded from the word "watch" and a photograph of a large, gold watch with a flexible, stretch strap etc.

When items vary only in size, colour, form or pattern the words used to describe such simple geometric stimuli are relatively unambiguous with no particularly strong
emotive overtones.

In English, form plays a dominant role in descriptions. Objects are often defined in terms of functional categories - for example, a cup is something to drink out of - but the function usually restricts the objects so used to a characteristic size and shape range. These two attributes, size and shape, are often used to distinguish between members of the same functional category - for example, cups and mugs. With simple, stylized material the form is even more dominant; one talks of "large red circles" and not "large circular reds" (the frame of reference, geometric figures, does not permit the latter description as a proper political one). Form is encoded as a noun qualified by adjectives describing the other attributes of the stimulus. Thus, form can be seen to be a different sort of attribute from size, colour, or pattern (this statement may not be true for people who do not speak an Indo-European language). Because form is different it was decided to vary only size, colour and pattern; these attributes are at least encoded in the same grammatical class although they may not be of equal importance to all subjects and nor are they free from constraints in the natural order of encoding.¹ There is one other reason for not varying form as an attribute; Archer (1962) suspected that sex differences occurred in his concept learning experiment because less women than men

¹ Lockhart & Martin (1969)
knew distinctive labels for the forms he used, squares and parallelograms. It seems likely that this difference would occur with New Zealand students.

Because of Heidbreder's (1946) work on concept formation it was decided not to vary number as an attribute. She found numbers to be at a higher level of abstraction than forms and, presumably, sizes, colours, and patterns.

The time of presentation was intended to be above threshold for reading words and for naming pictures. Given a constant presentation time, the time available for rehearsal would depend on the time taken to encode the stimuli and so it could not be controlled independently. Nonetheless, it could be measured, even if crudely, by a questionnaire, enabling some comparison of the amount of rehearsal for verbal and for pictorial material to be made.

Ducharme & Fraise (1965), Paivio & Yarmey (1966), and Corsini et al. (1969) suggested that the type of response required might affect recall scores as it might be more difficult to decode an image into an appropriate verbal response than to recall a word. They did not consider that the form of response required might affect the form of encoding itself; Ss might not encode pictorial material verbally if a pictorial recognition response was required whereas they probably would encode verbally if a verbal response was required. Ducharme & Fraise (1965) note
that, for adults, presentation of familiar pictorial material almost automatically evokes naming but it still seems worthwhile to investigate this experimentally by varying the type of response required.

It did not seem practical to use Bregman & Chamber's (1966) reconstruction method to investigate the effect of the type of response. They used seven levels per attribute but for reasons discussed later only three levels were to be used in this experiment. It was logically possible in their study to reconstruct all seven stimuli from a knowledge of only six but there was no evidence that Ss realized this. With three levels per attribute it seemed likely that some Ss would realize that they need remember just two stimuli in order to reconstruct all three.

Visual recall of sizes, colours, and patterns would require Ss to draw the stimuli they had seen; this would appear to be a more difficult and certainly a less familiar task than verbal recall. Also, it would be much more difficult for the experimenter to mark. Kaplan et al. (1966) did not say how they scored Ss' reproductions of line drawings. Thus, recognition appears to be the best method. Use of this method, however, always poses the problem of what the distractor items are to be. It seemed simplest to construct a set of stimuli comprising all the possible combinations of all levels on the three attributes
of size, colour, and pattern. On each trial two or three stimuli would be presented and then Ss would have to mark off the ones they had just seen on a sheet containing the complete set. Burdock (1963) points out that this is essentially a test of list membership in contrast to "yes/no" or multiple choice recognition tasks. S does not have to respond to each item; he may search the response sheet until he finds the stimuli he recalls having just seen and not even bother to look at a number of the other items. If the set of response items is constant over trials for both content and position, then it seems most probable that this search strategy would be employed. This means that the task is intermediate between free recall and the other recognition tasks - S may recall an item, search and find it, mark it, having checked that it does look like an "old" item, and then proceed to scan systematically until he finds two other items which he thinks are old (assuming that he saw three stimulus items).

One other factor seemed worth investigating as well as the verbal/pictorial stimulus and response factors. This was the compact/distributed factor used by Shepard, Hovland, & Jenkins (1961) in a series of concept learning experiments. They used pictorial material and had three levels of distributedness. Compact stimuli carried all information on one figure - for example, a triangle which
was large and black; distributed (1) stimuli had as many figures as there were attributes, but each figure varied in the same way; and distributed (2) stimuli were like distributed (1) stimuli except that each figure varied in a different way. It was decided to use compact and distributed (2) stimuli as these could be verbally encoded in the same words and for the rest of this thesis distributed (2) stimuli are referred to just as "distributed". Verbal versions were prepared but it is acknowledged that compactness and distributedness might not be the same for verbal and pictorial stimuli. In fact, this compact/distributedness factor was introduced because the physical spacing attributes in the words, "large red dotted", is wider than that in one blob which is large, red, and dotted. It was hoped that varying the distributedness of both verbal and pictorial stimuli might show if some such difference in the spacing of information was responsible for differences in memory for verbal and pictorial material.

4.1.2 Experimental Variables

To summarise, there were three main variables in this experiment and a fourth control factor:

i) The verbal/pictorial type of stimuli.

ii) The verbal/pictorial type of recognition task.

iii) The compact/distributedness of stimuli.

iv) sex.
4.1.3 Predictions

It was predicted that distributed stimuli would be more difficult to remember than compact ones, although the extent of the difference might not be the same for verbal and pictorial stimuli. The relationship between the verbal/pictorial stimulus and response factors was more complex. Consider the following diagram:

<table>
<thead>
<tr>
<th>Verbal Stimuli</th>
<th>Pictorial Stimuli</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal Responses</td>
<td>V → V</td>
</tr>
<tr>
<td>Pictorial Responses</td>
<td>V → P</td>
</tr>
</tbody>
</table>

If the encoding is the same, regardless of the type of stimulus input, and if the time taken to make an extra transformation from, say, pictorial to verbal encoding does not matter, then:

\[
\text{mean (V → V)} = \text{mean (P → V)}
\]

and

\[
\text{mean (V → P)} = \text{mean (P → P)}
\]

where the mean for each cell is the mean number of correct recognitions.

Similarly if the encoding is the same, regardless of the type of recognition output required, and if the time taken to transform from the encoding to the type of recognition
response does not matter, then:
\[ \text{mean} (V \rightarrow V) = \text{mean} (V \rightarrow P) \]
and \[ \text{mean} (P \rightarrow P) = \text{mean} (P \rightarrow V) \]

Thus, if encoding is independent of both the type of stimulus input and the type of response output required, and transformations do not affect the amount remembered, there will be no difference between the four cells. Note, however, that such a situation would not reveal what the encoding was — it could always be verbal, or non-verbal, or both. Also, different relationships among the cell means can be predicted from assumptions about how transformations at presentation or testing adversely affect recognition scores; still further relationships can be derived by assuming that verbal material is stored only verbally but that pictorial material is stored both verbally and visually.

This listing of possible predictions makes it clear that there was a need for more information on encoding than that provided by the recognition scores. Variations in visual and auditory similarity could not be carried out within the scope of this experiment and nor would they be easy to make with the sort of stimuli used. The only way to obtain the desired information was by means of a questionnaire. Use of such an instrument does not imply an assumption that Ss are always aware of how they
cope with tasks or that they can always verbalize what they are aware of; it merely implies that they may be able to do these things and that the responses so produced may relate reliably and meaningfully to the recognition scores and to the information on encoding obtained in other studies. This method is more similar to what Hayes (1968) calls "process tracing" than to classical introspection. Questionnaires of varying lengths have been used in experiments on perception, memory, and learning (e.g. Harris & Haber, 1963; Sheehan & Neisser, 1969; Paivio, Yuille, & Smythe, 1966; Montague, Adams, & Keiss, 1966) and Richardson (1965) and Dulany (1968) advocate that they should be used much more extensively.

In summary, this experiment investigated the effects on recognition scores of the type and spacing of stimulus information and the type of recognition response required; these effects, or the lack of them, were to be related to differences in encoding as reported in a questionnaire.

4.2 Methodological Considerations

4.2.1 Stimuli

The stimuli used varied on three attributes - size, colour, and pattern. There were three levels on each attribute - large, medium, and small; red, blue, and green; and dotted, striped, and checked. It is possible that the patterns were less readily and less unanimously
named than the sizes and the colours. Three levels were chosen because they gave 27 stimuli \((3^3)\), the maximum which could be readily fitted on a sheet of foolscap for the recognition trials. These 27 stimuli formed 108 pairs in which each member differed from the other on each attribute, as did the pairs used by Harris & Haber (1963). It was possible to present all 108 pairs in a 1 hour session; Harris & Haber could present only a subset of their possible pairs because they had more levels per attribute. Examples of the verbal and pictorial, compact and distributed stimuli are in Appendix I.

4.2.2 Response Sheets

Because of the expense of duplicating numerous coloured response sheets it was decided to use response templates which were fairly durable. Each subject had one cardboard template through which he marked his responses for each trial on a clean sheet of blank newsprint paper. Examples of a verbal and a pictorial response template are in Appendix I.

4.2.3 Apparatus

A Leitz (Wetzlar) 500 watt projector with a mechanical Alphax shutter was used. The calibration on this shutter was later checked with an integrating photometer which revealed considerable inaccuracy in calibration. Presentation times of 1 second and 0.5 seconds were chosen; these were subsequently found to be 0.7 seconds and 0.25 seconds respectively.
4.2.4 Subjects

All of the subjects were Stage I, II, or III Psychology students who were fulfilling course requirements, except for four non-Psychology students who were used only in the pilot studies or in the control group.

4.2.5 Questionnaire

The questionnaire was designed to provide information on encoding, general imagery, strategies, rehearsal, expectation of success, and census data such as age, year at university, and faculty. There was an open question, "Write briefly on how you tried to cope with the task in this experiment". This came before any of the specific questions on encoding and imagery. Most of the questions required "Almost, Often, Sometimes, Never" answers. A copy of the questionnaire is included in Appendix II.

4.3 Pilot Studies

These were undertaken to find the conditions suitable for the main experiment and to indicate the number of subjects needed.

4.3.1 Pilot Study 1

This study was carried out to find the error rate when Ss had to remember two slides, each presented for 0.7 seconds, and to check that the questionnaire could be understood.
Stimuli & Responses: Verbal compact stimuli and verbal response sheets, as described above, were used. Two stimuli differing on all three attributes were presented consecutively on each trial. As there were 108 such pairs, 108 trials were used. Each stimulus (i.e. a slide) appeared first four times and second four times. There were two practice trials using slides with the same sizes as for the other trials but with different colours and patterns and response templates to match. (See Appendix I). It was necessary to have practice trials so that the Pictorial Compact Stimulus Group could learn what the three sizes were. Therefore, all groups had to have these practice trials.

Timing: Presentation time was 0.7 ± .02 seconds. The time between presentation of the 1st slide and presentation of the 2nd slide was that required to push the slide holder across - approximately 0.2 seconds. The response time was 10-15 seconds per trial.

Instructions: These are in Appendix II. After an informal trial with one female subject, three males and three females were tested singly or in pairs. They all filled in a questionnaire after the experimental trial; a copy of this questionnaire is in Appendix II.

Results: Two error measures were used, the number of attributes wrong and the number of slides wrong. A slide
was counted as wrong if one or more attributes were wrong. Student's t-tests revealed no differences between males and females on either measure.

Mean number of attribute errors = 68.7 out of 648

\[ = 10.6\% \quad \text{S.D.} = 5.1\% \]

Mean number of slide errors = 47.2 out of 216

\[ = 21.3\% \quad \text{S.D.} = 8.6\% \]

The correlation between these two error measures was \( r = .99 \).

There appeared to be no comprehension problems with the questionnaire and some small spread of responses was obtained on each question.

Conclusions: The task appeared to be much too easy to be sensitive to experimentally manipulated variables. Therefore, it was decided to carry out a second pilot study with a lower presentation time per item.

4.3.2 Pilot Study 2

This was identical to Study 1 except that each stimulus was presented for about 0.25 seconds instead of 0.7 of a second. In spite of this brief presentation time no subjects reported difficulties in perceiving the slides; this is probably because they were familiar with all the possible words as these were on the response templates which they had studied for 1-2 minutes. Also, there were only 9 possible words.

Results: These were little different from those obtained in
the first pilot study. Values for the first pilot study are given in brackets.

Mean attribute error rate = 13.0% (10.6%)

S.D. = 12.7% (5.1%)

Mean slide error rate = 24.3% (21.8%)

S.D. = 13.6% (6.6%)

The correlation between the two error measures was $r = 0.93$.

Comparison with Pilot Study 1: Because the two error measures correlated so highly, a two way analysis of variance with sex and presentation time as the factors was carried out only for the number of attribute errors. This measure seemed more closely related to the amount Ss actually remembered than the number of slide errors.

The analysis of variance showed that there was no significant effect of exposure time or sex and nor was there an interaction between these two variables.

Conclusion: Because there was no significant increase in the error rate the 0.25 seconds exposure situation was no more satisfactory than the 0.7 seconds condition. In fact, the latter was felt to be preferable as subjects seemed to find the 0.25 seconds condition rather too fast. Also, there was no guarantee that the other types of stimulus material would be clearly perceived at the fast rate of presentation.
Therefore, it was decided to increase the number of slides per trial to three or four. This meant that some attribute levels would have to be repeated during one trial as if all three stimuli differed on each attribute then the third would always be predictable from the first two. Increasing the number of levels or the number of attributes to four would increase the number of possible stimuli to 64 \((4^3)\) or 81 \((3^4)\) respectively. These numbers are well above the number of response alternatives which can be fitted comfortably on to a foolscap response template so that it was most convenient to increase only the number of stimuli per trial to make the memory task more difficult.

4.3.3 Pilot Study 3

Two subjects, a male and a female, were given a brief testing session with four, and then three stimuli per trial. They found it very difficult to remember four stimuli per trial but three stimuli per trial seemed to provide a task with about the right level of difficulty. Therefore six subjects, three male and three female, were run to make sure that the three-stimuli task was significantly more difficult than the two-stimuli task. Presentation time was again 0.7 seconds.

Four subjects were discarded and replacement subjects were run. It was discovered that three subjects, all female, had devised a cunning way round the need to remember the three slides they had just seen - when they
saw a slide they placed a finger in the response template hole corresponding to that slide. Then, when all three slides had been presented, they merely had to remember the order in which they had seen the slides and write "1", "2", or "3" in the holes they had covered with their fingers. In consequence, all subsequent subjects were asked not to do this, nor to use any other such strategy; they were also told to watch the screen until they had seen all three slides. These changes to the instructions are included in Appendix II. The fourth subject was discarded because of a mistake in the order of presentation of the stimuli which made his responses impossible to mark.

The trial series was constructed by randomizing 54 trios of slides. These trios satisfied the following conditions:

1) the adjacent pairs of slides within the trio must differ on every attribute
2) the first and last slides must be as different as possible, given that
3) each slide will occur twice in each position.

The 54 trios were arrived at by trial and error as there seemed no formal way of producing trios to meet all of the criteria. The 54 chosen were a subset of the 1728 (27 x 8 x 8) possible trios satisfying condition 1).

Response time was increased to 25 - 30 seconds to allow
Ss to make three instead of two responses.

Results: Error measures were converted to percentage errors so that Pilot Studies 1 and 3 could be compared. As the correlations between attribute and slide errors were \( r = 0.99 \) for Pilot Study 1 and \( r = 0.97 \) for Pilot Study 3, only attribute errors were analysed in the comparison of the two studies.

<table>
<thead>
<tr>
<th>Table 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean percentage of attribute errors</td>
</tr>
<tr>
<td>2 slides/trial</td>
</tr>
<tr>
<td>Males</td>
</tr>
<tr>
<td>Females</td>
</tr>
</tbody>
</table>

An analysis of variance showed that three slides per trial were significantly more difficult to remember than two slides per trial (\( F (1,8) = 11.2, 0.01 < p < 0.05 \)). It does not matter whether this is due to the increase in the number of slides per trial, changes in response time and the number of trials, or the introduction of repeated attribute levels within a trial; an experimental task with a more suitable error rate has been found.

Conclusion: A suitable task for the main experiment required Ss to remember three slides on each trial. However, before beginning this experiment it seemed
advisable to compare the results for Pilot Study 3 with those from one of the other experimental conditions so that some estimate of the required number of Ss per cell could be obtained. Also, certain ambiguities had been found in the questionnaire and further testing was needed.

4.3.4 Pilot Study 4

This was identical to Pilot Study 3 except for the stimulus material. All of the previous pilot studies used verbal compact stimuli and verbal response sheets (V_{CV}); in this pilot study pictorial distributed stimuli and verbal response sheets were used (P_{DV}). This condition was chosen as it was expected to provide the most problems with the instructions. It also allowed comparison with V_{CV} on both the type and distributedness of the stimulus material. The correlation between slide and attribute errors was \( r = 0.99 \) for \( P_{DV} \) and \( r = 0.97 \) for \( V_{CV} \) so only attribute errors were analyzed.

Results: Table 2

<table>
<thead>
<tr>
<th></th>
<th>V_{CV}</th>
<th>P_{DV}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Males</td>
<td>153.3</td>
<td>120.7</td>
</tr>
<tr>
<td>Females</td>
<td>88.3</td>
<td>134.7</td>
</tr>
</tbody>
</table>
An analysis of variance showed that neither the experimental condition nor sex had any significant effect. Nor was the interaction significant. Using the necessary assumptions that neither cell means nor variances would vary over the range of sample sizes considered, critical sample sizes were calculated for i) a significant stimulus condition effect; ii) a significant sex effect; and iii) a significant interaction. The required sample sizes per cell were found to be 19, 6, and 4 respectively.

Bearing in mind that the total design was a $2 \times 2 \times 2 \times 2$ one with 16 cells, a sample size per cell of 6 seemed suitable. This meant that there would be 48 Ss at any one level of a factor. Thus, even if no verbal/pictorial or compact/distributed main effects emerged, any interactions with sex or any sex main effect would be likely to reach significance. Also, a total number of 96 subjects was suitable in terms of the number of subjects and the length of time for which they would still be available.

It was decided to tell subjects that a) they were not expected to remember everything so that when in doubt they should guess instead of doing nothing or telling the experimenter how they just could not cope; b) although the timing would be fairly strict, extra response time would be allowed for mishaps such as dropping pencils or
response sheets - the latter had to be extracted from under the response template after each trial and placed on a separate pile; and iii) they should work independently of one another, not worrying if they took longer to respond than the other subjects as long as they themselves worked within the time limit. It seemed that slow subjects worked well within the time limit when with others but tended to dawdle when alone. Therefore, it was decided to have at least two subjects for all experimental sessions in the main experiment. The maximum number who could be run together was three.

Some parts of the questionnaire were found to be ambiguous and additional instructions were given verbally when Ss were answering it. These additional instructions are in Appendix II.

4.3.5 Summary

From the results of four pilot studies it was decided to use three stimuli per trial in the main experiment with a presentation time of 0.7 seconds per stimulus. Six subjects per cell were required for this experiment. Necessary changes were made in the instructions for the experimental trials and the questionnaire.

4.4 Experiment 1

4.4.1 Design

The design was a 2 x 2 x 2 x 2 fully factorial one. The factors were the compact/distributedness of stimuli,
the verbal/pictorial type of stimuli or response templates, and the control factor of sex. One completely separate control group was run simultaneously to check that presentation time was well above threshold for all types of stimulus material.

4.4.2 Control Group

24 Ss were run to check that all types of stimulus material were clearly visible. There were three males and three females in each of the four conditions - verbal compact, verbal distributed, pictorial compact, and pictorial distributed types of stimulus material. Ss were informed that they were in a control group and that they were expected to find the task easy. After explanations on the type of slides they would see, presentation of the usual practice stimuli, and study of a corresponding response template, Ss were given 54 trials, each of the 27 slides being shown twice in a random order. On each trial one slide was shown for 0.7 seconds and then Ss wrote down the size, colour, and pattern. Three subjects were omitted, two from the pictorial distributed group and one from the pictorial compact group, because they had misunderstood instructions and could not label the sizes correctly. This was unlikely to happen in the main experiment where Ss had the response template, which showed that there were only three possible sizes, in front of them all the time.
Results:

Table 3

<table>
<thead>
<tr>
<th></th>
<th>Compact Stimuli</th>
<th>Distributed Stimuli</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal Stimuli</td>
<td>0.36%</td>
<td>0.82%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.59%</td>
</tr>
<tr>
<td>Pictorial Stimuli</td>
<td>1.34%</td>
<td>0.93%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.13%</td>
</tr>
<tr>
<td></td>
<td>0.85%</td>
<td>0.87%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.86%</td>
</tr>
</tbody>
</table>

4.4.3 Experimental Groups - Procedure

Two or three subjects were run together. Adjacent subjects always had different types of response templates. After two practice trials, 54 experimental trials were run. On each trial, three slides were presented consecutively for 0.7 seconds each with about 0.5 seconds between each slide, this being the time required to manually push the slide holder across and change a slide. Ss responded by marking "1" for the 1st slide, "2" for the 2nd slide, and "3" for the 3rd slide on the blank sheet underneath the response template. After the experimental trials Ss filled out the questionnaire on encoding and on general imagery. Copies of the instructions and the questionnaire are in Appendix II.

4.5 Results for the Memory Task.

The various error measures and the problem of incomplete responses are discussed. Then analyses of variance are reported.
4.5.1 Error measures: for each S there were ten error measures. These were obtained from the raw data by computer analysis.

\[ T = \text{number of slide errors (a slide was counted as wrong if one or more of the three attributes was wrong or if it was given the wrong series position number e.g. if it was called "2" when it had actually been the first slide),} \]

\[ TP = \text{number of slide errors except that the series position number did not matter,} \]

\[ A = \text{number of attribute errors,} \]

\[ S = \text{number of size errors,} \]

\[ C = \text{number of colour errors,} \]

\[ P = \text{number of pattern errors,} \]

\[ S + C + P = A. \]

\[ AP, SP, CP, \text{ and } FP \text{ bear the same relation to } A, S, C, \text{ and } P \text{ respectively as } TP \text{ does to } T, \text{ that is the responses were marked without regard for the reported position in the presentation trio.} \]

In the pilot studies A and T were correlated as a crude way of showing that only one need be analysed. The method was crude in that it was somewhat invalid as there were certain restrictive boundary conditions on each pair of measures e.g.

\[ 0 \leq T \leq A \leq 3T \leq 3n \text{ where } n = \text{number of slides presented over all trials.} \]
For the relations A & T, A & AP, T & TP, S & SP, C & CP, P & PP, $X^2$ tests on the bounded zones led to rejection of the hypothesis of equidense packing over the total possible areas. This implied a strong functional relationship between all the above pairs of measures. Therefore, it was decided to use only A, and S, C, and P in the analyses. (See Appendix III for details of the relationships.)

It is interesting to note that Ss seldom made transposition errors with correctly recalled slides. The number of transpositions was $T - TP$. As Conrad (1965) and Murdock & vom Saal (1967) have found, transpositions occur infrequently. A more detailed analysis of slide transpositions is difficult here as i) recall of a slide was actually recall of three attribute levels and not just recall of one letter or one word; and ii) the range of possible responses was so limited that successful guessing probably contaminates any measures of the number of transpositions. Over all groups the mean probability of a transposition was 0.036.

4.5.2 Incomplete responses: 4 Ss missed some responses. 2 of them missed only 1 out of 162 and another missed 6; their missing responses were filled in using a random number table. They came from the PDP female group and the PDV male group. The same procedure was followed for the 4 Ss in the pilot studies who missed 1 to 8 responses.
The 4th S, a VCV male, must have somewhat misunderstood the instructions as he missed 49 out of 162 responses. No more Ss were available at the time so his missing responses were generated from a random number table so that the T, TP, A, and AP scores for those responses corresponded to the means for those error rates over all the other 35 Ss. This was undertaken as a more conservative strategy than simply generating random responses.

4.5.3 Analyses of Variance: The main problem with the analysis of variance was whether or not compact/distributedness could be considered as a factor separate from the verbal/pictorial stimulus factor. Appendix IV discusses this in detail. It was decided to treat compact/distributedness as a factor on its own in the absence of any apparent interactions with the verbal/pictorial stimulus factor.

As a check, a one way analysis of variance was done on the four stimulus conditions. The verbal compact condition was found to be equivalent to the pictorial compact condition (Inverse F (92,1) = 128.2, 0.10 > p > 0.05) and the verbal distributed condition was equivalent to the pictorial distributed condition (Inverse F (92,1) = 14.6 p = 0.20). There is some disagreement over the interpretation of inverse Fs. On a Bayesian approach using likelihood functions, anything more significant than p = 0.20 seems acceptable. (R.A.M. Gregson, 1968, personal
communication). These results could not have been obtained had there been a significant verbal/pictorial stimulus type effect and so such a simple analysis cannot generally be used to solve problems of this kind.

The $2 \times 2 \times 2 \times 2$ analysis of variance on the number of attribute errors, $A$, was carried out using an adapted form of the IBM Scientific Subroutine Sample Program ANOVA. The program was altered to incorporate Box & Cox's (1964) procedure for finding the best transformation; this is defined as the transformation for which the sum of the error and interaction terms is at a minimum. The specific transformation is obtained from a search over a range of $\lambda$, the incremented power parameter in the equation

$$ y(\lambda) = \begin{cases} \frac{y^{\lambda-1}}{\lambda} & (\lambda \neq 0) \\ \log y & (\lambda = 0) \end{cases} $$

In the actual analysis the standardized variate $z(\lambda)$ is used.

$$ z(\lambda) = \frac{y^{\lambda-1}}{\lambda \hat{y}} $$

where $\hat{y} = \sqrt[n]{\frac{\sum_{i=1}^{n} y_i}{n}}$ i.e. the geometric mean of the untransformed data.

The posterior density assuming a uniform prior is

$$ p_u(\lambda) = k \{S(\lambda; z)\}^{-\frac{1}{2}} $$
$S(\lambda; z)$ is the sum of squares for the residual error and interactions of $z(\lambda); \nu_r$ is the degrees of freedom for the same.

$p_u(\lambda)$ is distributed approximately normally and so confidence limits can be found for the best $\lambda$. The program did not calculate these limits and, in fact, worked directly on $S(\lambda; z)$. This meant that a constant did not have to be obtained for each set of data so that

$$p_u(\lambda) d\lambda = 1.0$$

Table 4 gives a summary of the analysis of variance for the untransformed data. Results from the transformed data were little different. F ratios of less than one and negative variance components are omitted unless they occur for a main effect. Variance components are calculated on a four-way extension of Vaughan & Corballis's (1969) equations for a three-way design. All factors were fixed. As recommended by Vaughan (personal communication, 1969) all negative variance components were treated as zero for an estimate of the total variance.

On the untransformed data the only significant main effect was that for the compact/distributed factor, error scores being higher for distributed stimuli ($F(1, 80) = 8.30, 0.005 < p < 0.01$). This was not a strong effect (variance
ratio = 6.7%) and nor was the only significant interaction 
\( F(1,30) = 4.59, 0.01 < p < 0.05, \) variance ratio = 3.3%).

This three-way interaction is difficult to interpret; if anything, it suggests that males find mixed modes (e.g. Verbal stimuli, Pictorial response templates) more difficult than single modes while females find the opposite.

The best transformation was at \( \lambda = 0.50 \) which was not surprising as the distribution of the number of attribute errors was negatively skewed. The improvement in sensitivity for the main factors was small \( F = 9.15 \) compared with \( F = 8.30 \) and the three-way interaction did not vanish.

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>P</th>
<th>VC</th>
<th>VR</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>270.0</td>
<td>0.10</td>
<td>&gt;.25</td>
<td>-24.1</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>7722.1</td>
<td>2.99</td>
<td>&gt;.05</td>
<td>53.5</td>
<td>0.018</td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td>21450.3</td>
<td>8.30</td>
<td>&lt;.01</td>
<td>196.5</td>
<td>0.067</td>
</tr>
<tr>
<td>D</td>
<td>1</td>
<td>4523.8</td>
<td>1.75</td>
<td>&gt;.10</td>
<td>23.2</td>
<td>0.007</td>
</tr>
<tr>
<td>AB</td>
<td>1</td>
<td>44.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AC</td>
<td>1</td>
<td>5.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AD</td>
<td>1</td>
<td>1863.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BC</td>
<td>1</td>
<td>36.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BD</td>
<td>1</td>
<td>364.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD</td>
<td>1</td>
<td>184.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABC</td>
<td>1</td>
<td>3.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABD</td>
<td>1</td>
<td>11859.3</td>
<td>4.59</td>
<td>&lt;.05</td>
<td>96.6</td>
<td>0.033</td>
</tr>
<tr>
<td>ACD</td>
<td>1</td>
<td>2044.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BCD</td>
<td>1</td>
<td>44.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABCD</td>
<td>1</td>
<td>742.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ERROR</td>
<td>80</td>
<td>2582.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 4**

Summary of Analysis of Variance for Expt. 1

\( A \) - Verbal/Pictorial Stimulus Type
\( B \) - Verbal/Pictorial Recognition Task
\( C \) - Compact/Distributedness
\( D \) - Sex
\( VC \) - variance component
\( VR \) - variance ratio
The mean number of attribute errors over all groups was 101.8 which was an error rate of 20.9%. As the mean error rate for the control group was 0.86% it is unlikely that the memory error rate was much confounded by perceptual errors. Nonetheless, if the measure of perceptual errors corresponds to the difficulty and the time taken for perception of each type of stimulus and each attribute, then one would expect less time for rehearsal and hence higher memory error scores for those stimuli or attributes for which perceptual errors are more frequent (Mackworth 1963, 1964). This was not so as a comparison of Tables 3 and 5 shows.

<table>
<thead>
<tr>
<th></th>
<th>Compact Stimuli</th>
<th>Distributed Stimuli</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal Stimuli</td>
<td>85.4</td>
<td>114.9</td>
</tr>
<tr>
<td>Pictorial Stimuli</td>
<td>88.3</td>
<td>118.3</td>
</tr>
<tr>
<td></td>
<td>86.9</td>
<td>116.6</td>
</tr>
</tbody>
</table>

As Murdock & Ogilvie (1968) point out, this kind of error data is basically binary data pooled in some way. For example, the size of the first slide is recalled correctly or wrongly — and can be scored 0 or 1. In the
analysis of attribute errors, there is pooling over trials, position within a trial, levels of similarity between slides on a trial, and the three attributes of size, colour, and pattern. Unlike the kind of design Murdock & Ogilvie recommend, in which each S occupies a unique cell, there are six Ss in each unique cell in this experiment.

Changes over trials were not of any theoretical interest and so pooling over trials was maintained throughout all the analyses. Also, the 54 trios of slides used for the 54 trials were placed in one random order with no natural subdivisions for blocking over groups of trials. As a check, scores were plotted for two groups, $V_C V$ Males and $P_P P$ Females, and no very marked changes over trials were observed.

Differences between attributes were of interest, although in this experiment attribute affects are confounded with order of encoding. To compare memory for size, colour, and pattern directly would require experimental groups who had to remember only one such attribute, each group having to remember a different attribute. To test for attribute effects, an analysis of variance was carried out as before but with the attribute error scores ($A$) decomposed into size ($S$), colour ($C$), and pattern ($P$) scores, these constituting a fifth factor. On the untransformed data the main
effect for attributes was highly significant
\(F (2,160) = 19.5, p<0.001\) and a two way interaction
with verbal/pictorial presentation was just significant
\(F (2,160) = 4.08, 0.01<p<0.05\). The best transformation
was again at \(\lambda = 0.50\), but while increasing the significant
Fs it did not alter the significance levels \(F = 22.8\) and
\(F = 4.73\) for the main effect and interaction respectively).

Table 6

Mean errors for each attribute

<table>
<thead>
<tr>
<th>Size(S)</th>
<th>Colour(C)</th>
<th>Pattern(P)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal Stimuli</td>
<td>26.9</td>
<td>33.2</td>
</tr>
<tr>
<td>Pictorial Stimuli</td>
<td>33.4</td>
<td>29.4</td>
</tr>
<tr>
<td></td>
<td>30.2</td>
<td>31.3</td>
</tr>
</tbody>
</table>

As can be seen from Table 6, \(S<C<P\). This does not
quite correspond to the rank order of attributes in the
perception task; that is, the control group condition.
There, the average percentage error rates were \(S = 0.77\%\),
\(C = 0.31\%\), and \(P = 1.43\%\).

The interaction, while significant only at the 0.05
level, is meaningful. For verbal stimuli, size was easier
than colour; for pictorial stimuli, colour was easier
than size. There may have been some mislabelling or
difficulty in labelling the size of pictorial stimuli. This explanation is supported by the control group results which showed that for verbal stimuli, \( S = 0.16\% \) & \( C = 0.62\% \), and for pictorial stimuli, \( S = 1.39\% \) & \( C = 0.0\% \). Alternatively, it may be the case that it is easier to remember actual colours rather than colour words.

It was thought advisable to see if there were any effects of position and similarity within trials, although such effects were not of theoretical importance. Interactions between attributes, positions, and similarity were also not of interest and so the decomposition of attribute errors into attribute, position, and similarity error scores was carried out separately, each variable being analysed as the 5th factor in the basic 2 x 2 x 2 x 2 experimental design.

Position had a very significant main effect \( (F (2,160) = 69.3, p<0.001 \) for the untransformed data; \( F = 72.6 \) for the best transformation with \( \lambda = 0.50 \)). As can be seen in Table 7, the first slide was the easiest to remember, then the last, with the middle slide being the most difficult. Two interactions were significant at the 0.01 level, one at the 0.05 level, and one just reached significance \( (p<0.05) \) for \( \lambda = 0.50 \). Note that although the best transformation is one for which the sum of the interaction and error terms is least, this does not imply
that all interactions will be less for the transformed than for the untransformed data. The three-way interaction (0.001 < p < 0.01) between position, compact/distributedness, and sex was not interpretable. The two way interaction between compact/distributedness and position indicated that distributedness adversely affected recall of only the first two slides. The barely significant interactions were between position and the type of response template (verbal or pictorial) and between position and sex.

The two interactions with compact/distributedness do not indicate that it is wrong to treat it as a separate factor; this would be implied only by interactions also involving the verbal/pictorial stimulus factor.

**Table 7**

<table>
<thead>
<tr>
<th>Mean number of attribute errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st Slide</td>
</tr>
<tr>
<td>Compact Stimuli</td>
</tr>
<tr>
<td>Distributed Stimuli</td>
</tr>
<tr>
<td>26.3</td>
</tr>
</tbody>
</table>

Similarity between the three slides presented on one trial was defined as the number of attribute levels shared by the first and third slides (the 1st and 2nd, and the 2nd and 3rd slides were completely different on all attributes.
so that common attribute levels could occur only with the 1st and 3rd slides). There were four levels of similarity possible - none, one, two, or three attribute levels in common. With three attributes the same, the 1st and 3rd slides were identical. As the trios of slides were constructed to be as dissimilar as possible (see Pilot Study 3), there were unequal numbers of trials at each level of similarity. Attribute error scores were converted to probabilities of error at each similarity level.

Table 8

Similarity and attribute errors

<table>
<thead>
<tr>
<th>Similarity Level</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of trials</td>
<td>7</td>
<td>18</td>
<td>27</td>
<td>2</td>
</tr>
<tr>
<td>Mean probability of an attribute error</td>
<td>0.21</td>
<td>0.25</td>
<td>0.19</td>
<td>0.13</td>
</tr>
</tbody>
</table>

Analysis of variance revealed a very significant main effect of similarity (F (3,240) = 33.72, p<0.001) and no interactions. The best transformation came at the odd value of \( \lambda = -2.20 \) and was the only transformation in all these analyses which markedly affected an already significant F (F = 33.72 for the untransformed data and 52.25 for \( \lambda = -2.20 \)). Because of the very unequal numbers of trials
at each similarity level, these results probably do not mean very much.

4.5.4 Summary of analysis of variance results: In the analysis of variance on attribute errors (A) only one of the four experimental factors produced a significant main effect. This was the compact/distributed factor - compact stimuli were remembered better than distributed stimuli. Error scores were not affected by the sex of the Ss or by the verbal or pictorial type of stimuli or response sheets.

The three attributes, size, colour, and pattern differed significantly and so did the three slide positions, 1st, 2nd, and 3rd within a trial. There was also a significant main effect from the level of similarity between slides on any one trial.

Several interactions were noted.

The best transformations of the data, as found by Box & Cox's (1964) method, increased the F ratios for the main effects but did not alter the pattern of results.

4.6 Questionnaire Results

A copy of the questionnaire is in Appendix II. The figures given underneath each question are the percentages of 96 Ss who responded with each category.

All questions were cross tabulated with all other questions and with error data using Russell's (1969) program.
In this section responses to the open question at the beginning of the questionnaire are discussed and then the relation between the actual and expected percentage of correctly recalled slides is reported. Section 3.6.3 shows that none of the experimental variables affected encoding. Finally, factor analysis results are presented. All questions except the open question and question 7 were used in the factor analysis. Question 7 was omitted because it originally applied to restrictions on pairs of stimuli and became unintelligible to Ss who were presented with trios of stimuli, even though additional instructions were given verbally by the experimenter (Appendix II).

4.6.1 Open Question

This question was included merely to see if Ss responded about matters which were not covered in the main part of the questionnaire. Answers related mainly to strategies, such as concentrating on the 1st and 3rd slides (mentioned by 11.5% of Ss), or to the usual response order (mentioned by 25.0% of Ss). No relationships between the several types of strategies were apparent.

Position coders were the 16.6% of Ss who reported that when they saw a slide they mentally translated it into its position on the response template. Position coding was found to occur more often with Ss who saw verbal stimuli \( (\chi^2 = 6.08, \text{df} = 1, 0.01 < p < 0.05) \) and was related to better
recall of size \( \chi^2 = 3.27, \, df = 2, \, 0.01 < p < 0.02 \). The latter result was not surprising as the main divisions on the response templates were in terms of size.

Ss who saw distributed stimuli were more likely to mention that they noted similarities between slides on any one trial \( \chi^2 = 6.06, \, df = 1, \, 0.01 < p < 0.05 \). There were no other significant relationships.

Replies indicated that any similar, subsequent experiments should have forced choice questions on strategies and response orders. With open questions one cannot tell if Ss actually did not employ some strategy or if they just failed to mention it.

4.6.2 Expected percentage of Successes

In question 6, Ss had to reply what percentage of the time they thought they remembered whole slides correctly. The response alternatives were "0%, 10%, ..., 90%, 100%". These expected percentages were regressed on:

i) the percentage of slides correct
\[ \left( \frac{[162 - T]}{162} \right) \times 100 \quad \text{see 4.5.1}, \]

ii) the percentage of slides correct, rounded to the nearest 10%
and iii) the percentage of slides correct, truncated to the nearest 10%.

For the regression on i) there was a significant linear relationship \( F (1, 94) = 54.6, \, p < 0.001 \) of the form
Expected $\% = 12.2 + 0.61$ (% correct)

There was also a significant 5th degree component

$F (1,90) = 4.76, 0.01 < p < 0.05$ but this was presumably
due to the greater accuracy for the percentage correct
which could vary in steps of less than 1 whereas the
expected percentage varied in steps of 10. Regressions
on both the rounded (ii) or the truncated (iii) percentages
confirmed this as they gave only significant linear relations-
ships ($F = 58.3$ and $52.3$ respectively). The IBM Scientific
Subroutine Sample Program POLRC, for polynomial regression,
was used.

These analyses show that Ss' estimates bear a close
relation to their scores but that Ss who do well tend to
under-estimate and those who do poorly tend to over-estimate
the percentage of times they were right.

4.6.3 Effects of the Experimental Variables

It was of theoretical importance to see if encoding was
affected by the experimental variables, particularly the
verbal/pictorial type of stimuli or responses. The questions
specifically on encoding were questions 2, 3, 4, and 5, on
verbal, auditory, form, and visual encoding respectively.
As Table 9 shows, none of the four experimental variables
affected responses to any of these questions.
Table 9
Experimental variables and encoding

<table>
<thead>
<tr>
<th></th>
<th>Q2</th>
<th></th>
<th>Q3</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\chi^2$</td>
<td>df</td>
<td>$p$</td>
<td>$\chi^2$</td>
</tr>
<tr>
<td>Verbal/Pictorial Stimuli</td>
<td>0.79</td>
<td>1</td>
<td>$0.30&lt;p&lt;0.50$</td>
<td>0.63</td>
</tr>
<tr>
<td>Verbal/Pictorial responses</td>
<td>4.28</td>
<td>2</td>
<td>$0.10&lt;p&lt;0.20$</td>
<td>1.23</td>
</tr>
<tr>
<td>Compact/ Distributedness</td>
<td>0.18</td>
<td>2</td>
<td>$0.90&lt;p&lt;0.95$</td>
<td>2.80</td>
</tr>
<tr>
<td>Male/female</td>
<td>2.78</td>
<td>2</td>
<td>$0.20&lt;p&lt;0.30$</td>
<td>0.81</td>
</tr>
<tr>
<td></td>
<td>Q4</td>
<td></td>
<td>Q5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\chi^2$</td>
<td>df</td>
<td>$p$</td>
<td>$\chi^2$</td>
</tr>
<tr>
<td>Verbal/pictorial stimuli</td>
<td>0.14</td>
<td>1</td>
<td>$0.70&lt;p&lt;0.80$</td>
<td>5.32</td>
</tr>
<tr>
<td>Verbal/pictorial responses</td>
<td>0.10</td>
<td>1</td>
<td>$0.90&lt;p&lt;0.95$</td>
<td>3.05</td>
</tr>
<tr>
<td>Compact/ Distributedness</td>
<td>0.14</td>
<td>1</td>
<td>$0.70&lt;p&lt;0.80$</td>
<td>2.60</td>
</tr>
<tr>
<td>Male/female</td>
<td>0.10</td>
<td>1</td>
<td>$0.90&lt;p&lt;0.95$</td>
<td>1.64</td>
</tr>
</tbody>
</table>

One possible explanation for the compact/distributed effect on recognition scores is that distributed stimuli take longer to perceive and hence there is less time to rehearse them. However, the relationship between compact/distributedness and Q8, on rehearsal, was nonsignificant ($\chi^2 = 3.00$, df = 2, $p > 0.20$). Thus, assuming that questionnaire responses constitute admissible evidence, differential
rehearsal cannot be adduced as an explanation of superior recognition scores for compact stimuli.

Actually, none of the experimental variables related to anything other than strategy and faculty. The relations with strategy are discussed in section 4.6.1. The relation with faculty was the obvious one that there are fewer girls than boys doing science! \( \chi^2 = 6.1, \text{df}=1, 0.01 < p < 0.05 \).

4.6.4 Factor Analysis

This was used merely as an exploratory, data reduction tool to help sort out relationships between experimental variables, errors, and answers to the questionnaire. Thus some of the data was interval (age, errors etc.), some was ordinal (answers to "always", "often", "sometimes", "never" questions), and some was dichotomous ("yes/no" questions and experimental conditions). Product moment correlations were used. Kendall & Stuart (1966, p.311) note that if tetrachoric or biserial coefficients are used the matrix of correlations is not necessarily positive definite and in certain cases the latent roots may turn out to be negative. There were 18 variables and 96 Ss.

The IBM Scientific Subroutine Sample Program, FACTO, was used. This program performs a principal components solution using the diagonalization method developed by Jacobi and adapted by Von Neuman (Greenstad, 1969). It then
performs a varimax rotation (Kaiser, 1959). A stopping rule of root-one was adopted for the number of factors. Horn (1966) points out that this yields an overestimate of the number of common factors. This is doubtless even more the case with this data because so little of it is interval.

Eight factors with eigenvalues of greater than one were extracted and rotated. The eigenvalues, percentage eigenvalues, and all rotated factor loadings greater than 0.30 are listed in Table 10. Here the use of loadings greater than 0.30 is probably, again, an overinclusive strategy because of the ordinal nature of most of the data. There is no general factor and the variance taken up by each factor flattens out after the first two factors and is virtually the same small value (6%) for factors VI, VII, & VIII.

Factor I appears to be inversely related to recognition error scores. The distributedness of stimuli has a negative loading on this factor which was to be expected from the analysis of variance results. Encoding in words (Q2) and "hearing" words (Q3) have positive loadings. Expectation of success (Q6) has a very high loading on this factor.

1. Perhaps maximum likelihood factor analysis would have been preferable but although copies of two such programs mentioned by Jöreskog & Lawley (1966) were available they could not be used without rewriting to make them compatible with the word length of the University's IBM 360/44 computer.
Table 10
Rotated Factor Loadings for Expt.1

<table>
<thead>
<tr>
<th>Variables</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>V</th>
<th>VI</th>
<th>VII</th>
<th>VIII</th>
<th>h^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>V/P Stimuli</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>87</td>
<td></td>
<td></td>
<td></td>
<td>79</td>
</tr>
<tr>
<td>V/P Resps.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-67</td>
<td></td>
<td></td>
<td></td>
<td>58</td>
</tr>
<tr>
<td>C/D</td>
<td>-45</td>
<td></td>
<td></td>
<td></td>
<td>-46</td>
<td></td>
<td></td>
<td></td>
<td>63</td>
</tr>
<tr>
<td>M/F</td>
<td></td>
<td>82</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>77</td>
</tr>
<tr>
<td>Age</td>
<td>-83</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>74</td>
</tr>
<tr>
<td>Year</td>
<td>-75</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>71</td>
</tr>
<tr>
<td>Faculty</td>
<td></td>
<td>-66</td>
<td></td>
<td></td>
<td>-31</td>
<td></td>
<td></td>
<td></td>
<td>69</td>
</tr>
<tr>
<td>Words Q2</td>
<td>46</td>
<td>49</td>
<td></td>
<td></td>
<td>-30</td>
<td></td>
<td></td>
<td></td>
<td>62</td>
</tr>
<tr>
<td>Hearing Q3</td>
<td>59</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>55</td>
</tr>
<tr>
<td>Form Word Q4</td>
<td></td>
<td></td>
<td>-56</td>
<td>37</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>56</td>
</tr>
<tr>
<td>Seeing Q5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>88</td>
<td></td>
<td></td>
<td></td>
<td>81</td>
</tr>
<tr>
<td>% Expect Q6</td>
<td>72</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>66</td>
<td></td>
<td></td>
<td>64</td>
</tr>
<tr>
<td>Rehearsal Q8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>71</td>
<td></td>
</tr>
<tr>
<td>General Imagery Q9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>74</td>
<td></td>
<td>62</td>
</tr>
<tr>
<td>Breakfast Imagery Q10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interest in Art Q11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>80</td>
<td></td>
<td></td>
<td>77</td>
</tr>
<tr>
<td>Activity in art Q12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>78</td>
<td></td>
<td></td>
<td>79</td>
</tr>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-84</td>
<td></td>
<td></td>
<td></td>
<td>73</td>
</tr>
</tbody>
</table>

Eigenvalues: 2.61 1.98 1.54 1.46 1.27 1.18 1.12 1.06

% e.v.: 14 11 09 08 07 06 06 06

Decimal points have been omitted except for the eigenvalues.

V/P = verbal/pictorial factor  C/D = compact/distributedness
S/A = silent/voicing factor    M/F = male/female factor
F/Part = full/part-time dichotomy  A = attribute errors
h^2 = communality  % e.v. = percentage eigenvalue
which agrees with the separate analysis of the relationship between expected and actual percentages correct in 4.6.2.

Factor II looks like an age factor and could be dismissed as irrelevant were there not a loading on it for Q2, the question on encoding in words. A comparison between age and Q2 gave a $\chi^2 = 11.5$ for df = 3 and $p<0.01$. Eighteen and nineteen year olds had a modal response of "always" while older students had a modal response of "often" for Q2. Possibly this is just a function of increased caution in the use of extremes, older students having thought and been taught more about the meaning of category words.

Factor III would appear to be theoretically irrelevant as only sex and faculty have loadings on this factor (c.f. 4.6.3).

Factor IV is an art factor and as it is not related to errors, encoding, or experimental conditions it is of no further interest. Likewise, factor V is a general imagery factor. The questions on interest in art and general imagery were included just to see if they related to the use of imagery in the specific experimental task; they did not.

Factors VI, VII, VIII should probably be ignored as they take up so little of the variance (c.f. Horn, 1966)

To summarize, the factor analysis revealed only one important and theoretically interesting factor. This was factor I which appeared to be a recognition memory factor.
The only new information contributed by this analysis was that recognition memory was better the more stimuli were verbally encoded (Q2) and "heard" (Q3). Recognition memory was also related to the compactness of stimuli and was closely related to Ss estimation of their success at the memory task.

4.7 Discussion

In this experiment, no differences were found between verbal and pictorial stimuli. This is in direct contrast to the body of findings discussed in the literature review (Chapter 3) which showed better recognition and recall for pictorial stimuli. This experiment is not a replication of any of these studies and so detailed comparisons are difficult. However, it would seem that the main difference between this experiment and the free recall and recognition studies is in the number of relevant attributes for each item and the attempt to equate for the number of attributes carried by each type of stimulus material. There is no a priori reason to suppose that having to remember three attributes on three items instead of, say, twenty single items (Lieberman & Culpepper, 1965) should differentially affect pictorial and verbal material. Thus, it would seem most likely that the absence of a verbal/pictorial stimulus effect was the result of equating for the number of attributes carried by each type of stimulus. Conversely, it could be said that results from this experiment suggest that
other findings are a direct consequence of failure to control for the number of attributes carried by verbal and pictorial stimuli.

The use of a questionnaire enables exploration of the single versus dual encoding explanation for the difference between memory for verbal and pictorial stimuli. In this experiment there were no differences in recognition scores for verbal and pictorial stimuli and nor were there any differences in the amount of verbal or visual encoding, as measured on a four-point scale. The type of stimuli or responses did not affect encoding for this task.

However, in the absence of any effect from the experimental variables, it was found that the more verbal and auditory encoding reported, the better the recognition scores. This was shown clearly by Factor 1 of the factor analysis. A one-way analysis of variance of the effect of auditory encoding (Q3) on recognition scores gave the same result ($F(3,92) = 5.35, p<0.01$) and an analysis of verbal encoding (Q2) showed a similar but non-significant trend ($F(2,93) = 2.89, 0.05<p<0.10$). These differences between subjects correspond to differences between groups found by Murray (1965a, b; 1966a, b, c; 1967a, b) and Murray & Roberts (1968) in a series of experiments in which the amount of voicing was manipulated experimentally. Using lists of consonants, they found that voicing had a beneficial
effect on recall, except with lists of high auditory similarity. Why some subjects should spontaneously verbally encode more than others is not known.

The level of visual encoding, on the contrary, was not related to recognition scores. It had no loadings greater than 0.09 on the first five of the rotated factors and a one-way analysis of variance of the effect of visual encoding (Q5) on recognition scores was nonsignificant ($F(3,92) = 1.07$, $p>0.25$). For this task it would appear that Neisser (1967, p.157) is right in regarding imagery as a cognitive luxury. Nonetheless, it was quite a popular one.

It is interesting to note the distribution of responses for the three encoding questions.

**Table 11**

The percentage of 96 Ss responding with each category

<table>
<thead>
<tr>
<th></th>
<th>Never</th>
<th>Sometimes</th>
<th>Often</th>
<th>Always</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q2 &quot;thinking&quot; in words</td>
<td>0.0</td>
<td>12.5</td>
<td>45.8</td>
<td>41.7</td>
</tr>
<tr>
<td>Q3 &quot;hearing&quot; words</td>
<td>7.3</td>
<td>18.7</td>
<td>34.4</td>
<td>39.6</td>
</tr>
<tr>
<td>Q5 &quot;seeing&quot; stimuli</td>
<td>21.9</td>
<td>40.6</td>
<td>27.1</td>
<td>10.4</td>
</tr>
</tbody>
</table>

Thus, all Ss verbally encoded to some extent although the proportion of those who "heard" their encodings was less than those who just "thought" in words. Visual encoding was
much less common, but was used at least to some extent by 78.1% of Ss. Table 12 shows that most Ss, even those in the verbal stimulus, verbal response group, did occasionally encode visually.

Table 12

Number of Ss in each experimental group who visually encoded.

<table>
<thead>
<tr>
<th></th>
<th>Never</th>
<th>Sometimes</th>
<th>Often</th>
<th>Always</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal Stim. &amp; Resps.</td>
<td>3</td>
<td>10</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>Verbal Stim. Pict. Resps.</td>
<td>6</td>
<td>7</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>Pict. Stim. Verbal Resps.</td>
<td>8</td>
<td>10</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Pict. Stim. &amp; Resps.</td>
<td>4</td>
<td>12</td>
<td>6</td>
<td>2</td>
</tr>
</tbody>
</table>

The relative likelihoods of all the different forms of encoding (listed at the end of section 2.4) in the diagrammatic model for visual input cannot be estimated from the results of this experiment because there was no question to distinguish between visualizing words and visualizing pictures of referents or associates. Comparison with the verbal encoding questions is of no help as there is no a priori reason to expect verbal and visual encoding to be mutually exclusive.

The absence of a relationship between the amount of visual encoding and error scores poses a problem - presumably other processes must be postulated so that the model of
encoding can be related to recognition scores. Some understanding of differences in decay, interference, rehearsal or retrieval for verbal and visual encoding is needed before the model for encoding becomes a model for memory. Contrary to Bartlett's (1932) results, Ss who reported visual imagery did not expect to do better than Ss without imagery ($\chi^2 = 1.4, df = 4, p > 0.80$).

Reasons for the compact/distributed effect are not clear. The pattern of perceptual errors was not the same as the pattern of memory errors. Nor was there any reported difference in rehearsal for the two groups. It might, however, be difficult for Ss to estimate how much they rehearse, particularly if rehearsal is not a series of discrete repetitions but a portion of a person's limited capacity, as Posner & Rossman (1965) suggest. Gorfein & Stone (1967) found that presenting trigrams for periods corresponding to their pronunciation latencies removed differences in recall between items which differ in ease of pronunciation. A similar effect might occur here, although latencies of pronunciation might not be a very suitable measure of the time needed for perception. In Experiment 2 (Chapter 5), where half the Ss had to vocalize at presentation, it was noticed by the experimenter that Ss tended to wait before reading out verbal compact stimuli and then to read the three words as a phrase. On the other hand, Ss in the
verbal distributed condition tended to process each word separately, reading one then another. For this reason, some measure of processing time such as latency plus reading time might be suitable but latency alone would be misleading.

There is an alternative to the "perception time" explanation of the compact/distributed effect. All the recognition templates showed compact stimuli (Appendix 1). Distributed stimuli may have been remembered as well as compact ones but the appropriate answers were of a different form from the stimuli. Thus the correct answers may not have been "recognized", or else the time taken to transform each remembered item into the required form might have delayed recall for the items recalled second and third on each trial.

Twenty seven distributed stimuli could not have been fitted on to foolscap templates and any other arrangement of attribute levels would have introduced other contaminating variables. For this reason, only compact response templates were used.

It may seem odd that rehearsal (Q8) did not appear in any of the first six factors. This could be taken as evidence that Ss are incapable of reporting on the amount of rehearsal which takes place. Possibly they do have difficulty in so doing but there actually was a relation between the amount of rehearsal and error scores. This relation was nonlinear and so resulted in a very low product
moment correlation between the two. When the "Often" rehearse Ss were compared with all the others they made significantly less errors ($t' = 3.04, df = 19, p<0.005$). Rehearsal takes time; those Ss who always rehearsed what they had seen may have occasionally missed the next slide while rehearsing the previous one.

The attribute, position, and similarity effects are of little theoretical interest except that, in conjunction with the open question (Q1), they suggest that even with a simple task such as the one used in this experiment strategies are adopted; various parts of the information to be remembered are assigned different priority values.

4.8 Summary

With a task in which Ss had to remember three sequentially presented stimuli, verbal encoding was predominant regardless of the type of stimuli or the type of recognition responses required. In contrast with most other studies, pictorial stimuli were not remembered better than verbal stimuli. This was thought to be because the number of attributes carried by each type of stimulus had not been equated. Visual encoding occurred equally often for verbal and pictorial stimuli and where verbal or pictorial recognition responses were required. Visual encoding was not related to recognition scores and appeared to be what Neisser (1967, p.157) called a "cognitive luxury".
CHAPTER 5

Experiment 2

5.1 Introduction

In Experiment 1 it was found that verbal encoding predominated regardless of the type of stimulus material or recognition response required. The task involved memory for a series of three items, each item having three attributes, so that nine attribute levels had to be remembered on each trial. Response order was free but the serial position of each item had to be reported. The sequential nature of speech suggests that verbal encoding would be particularly suitable for a series of items; this was so in Experiment 1.

Perhaps imagery would be more important in nonsequential tasks such as memory for one item. Dominowski & Gadlin's (1968) third experiment (3.2) was not a satisfactory test of this hypothesis as they did not control for the number of attributes carried by verbal and pictorial stimuli and the distinction between perceptual and memory errors was not clear. Also, because they were primarily interested in paired-associate learning, they presented word-number and picture-number pairs for recall instead of just words or pictures. Therefore it was decided to carry out a second
experiment using the same stimulus material as in Experiment 1 but with a nonsequential task.  

The Peterson & Peterson (1959) technique, also used by Dominowski & Gadlin, seemed the most suitable as items could be presented singly and recalled after a certain period of interpolated task activity. However, it seemed advisable to ensure perceptual accuracy by using a presentation time longer than Dominowski & Gadlin's 0.3 seconds.  

A variety of interpolated tasks have been used such as naming colours (Wickens, Born, & Allen, 1963; Dominowski & Gadlin, 1969), reading six digit numbers (Loess & Waugh, 1967), reading three digit numbers (Talland, 1967), and counting backwards in threes (e.g. Peterson & Peterson, 1959; Murdock, 1961; Talland, 1967). Murdock (1965) has suggested the use of a concurrent subsidiary task like card sorting and Posner & Rossman (1965) have also investigated the effect of the difficulty of the interpolated task upon recall. La Berge & Winokur's (1965) shadowing technique is perhaps the most effective method of preventing rehearsal but it seems unlikely that complete prevention will be accomplished for all Ss with any method. The best which can be hoped for is some measure of the activity on the interpolated task. For 

1. This decision was made some months before Paivio & Csapo’s important article comparing recall of concrete and abstract words and pictures in sequential and nonsequential tasks was published in May, 1969, and even before Paivio, Yuille, & Roger's (March, 1969) article on noun imagery and meaningfulness in free and serial recall was available at the University of Canterbury.
instance, Wickens & Simpson (1968) found that differential amounts of backwards counting accounted for the facilitary effect of high motivation. However, Wickens (personal communication, 1969) states that in other experiments not involving motivation levels the amount of counting has been found to be constant over groups. Thus, although counting backwards in threes does not completely prevent covert rehearsal (Groninger, 1966) it is still a useful method. It is also one of the simplest to present and so was used in this second experiment.

In Experiment 1 Ss who said that they "heard" words did better than those who did not, which was in agreement with the studies by Murray (1965a, b; 1966a, b; 1967a, b) and Murray & Roberts (1968) on experimental manipulation of amounts of voicing. They used lists of consonants—does voicing have a facilitative effect when there is only one item to remember? The size, colour, and pattern on each slide are not likely to be confused with one another so that if articulation serves to make items more distinctive, as Murray suggests, then it should have rather less effect with only one three-attribute item to be recalled. The most likely effect would be a reduction in proactive inhibition (FJ) over trials. Peterson & Peterson (1959) did compare silent versus vocal rehearsal during presentation and found no effect but they used auditory presentation which
has been shown to give results different from those for visual presentation (Mascet, 1967). For that reason it still seemed worthwhile to introduce a silence versus voicing at presentation factor into this second experiment.

If the attribute differences obtained in Experiment 1 were due to the different importance attached to them (mentioned spontaneously by 9.4% of Ss), then having to voice all three attributes on each trial might reduce differences between attributes without altering the total amount recalled. Voicing might also reduce the compact/distributed effect although presentation time is probably the main factor involved in that.

There was no verbal/pictorial response effect in Experiment 1 and the response effects found by Dilley & Paivio (1969) and Ducharme & Fraisse (1965) occurred with children. Accordingly it seemed reasonable to use only verbal recall in Experiment 2. If adult Ss were familiar with the stimulus material before the recall trials then they should have little difficulty with written responses, even if these involved decoding a pictorial image, as the appropriate verbal code should be readily available. The only effect expected from requiring verbal responses might be to raise the overall level of verbal encoding.

As always, there was a need to ensure that perceptual errors were not interpreted as memory errors. Therefore it
was decided to duplicate Experiment 2 with a 0 seconds retention interval as a check on perceptual errors.

5.2 Experimental Variables

The experimental variables for Experiment 2 were the same as those for Experiment 1 except that the verbal/pictorial response factor was replaced by a silence/voicing factor. The control for perceptual errors was incorporated into the total design as a fifth factor instead of being run as a separate control group. The five factors were:

i) verbal/pictorial type of stimuli
ii) compact/distributedness of stimuli
iii) silence/voicing at presentation
iv) sex
v) retention interval - 0 or 30 seconds.

Stimuli were the same as in Experiment 1 (see Appendix I) and the Peterson & Peterson (1959) method of item presentation, backwards counting, and recall was used. A questionnaire on encoding at presentation and imagery at recall was given to subjects who were Stage I or Stage II Psychology students.

5.3 Pilot Study

Several informal sessions using verbal or pictorial stimuli showed that Ss were remembering about 65% of the slides and 85% of the attributes correctly at a retention interval of 30 seconds. Therefore, it was decided to carry out a pilot study comparing recall after 30 and 60 seconds.
Twenty four Ss were used, six males and six females at each retention interval.

5.3.1 Method

Stimuli: To enable comparison with results from Melton's (1963b) study, verbal compact stimuli were used. Ss had to read the words aloud at presentation. Practice stimuli carried different attribute levels for colour and pattern from those on the experimental slides. (Appendix I).

Timing: So that Ss were quite familiar with the type of stimulus material, the practice slide viewings lasted about 5 seconds each. On every trial, the slide to be remembered was presented for 0.8 seconds and then, after an inter-slide interval of one second, the number slide was presented for 0.7 seconds. Two projectors, both Leitz (Wetzlar), were used and the presentation times were the maximum possible using mechanical Aphan shutters which purported to go up to 1 second. The inter-slide interval was timed electronically on equipment with an accuracy of ± 1°, built by the technical staff of the University of Canterbury Psychology Department. Sperling's (1963) work suggests that the slides would be perceptable in iconic memory for longer than the presentation times as the room was just dimly illuminated by a "natural light" bulb over S's response sheet.

Procedure: Each 3 was run individually and had two 2 minute periods of practice at backwards counting, 3 slide viewings
and one practice trial, and 27 experimental trials. Thus, each S saw all stimulus slides. Three trial orders were used; these were constructed so that, summing over orders, each attribute level was presented on each trial. Two orders of numbers for backwards counting were also used. Over the 27 trials all nine digits, excluding zero, were used equally often as the last digit in the number from which to count backwards. Wickens (personal communication, 1969) notes that it is probably the last digit which most affects the initial subtractions. On each trial Ss attempted to count backwards to a 2 second beat. A copy of the instructions is in Appendix V.

5.3.2 Results

Attribute errors were used in both analyses.

i) An analysis of variance on errors summed over all trials showed neither retention interval, sex, nor their interaction to be significant ($F (1,20) = 0.33, 0.38, \& 0.27$ respectively).

ii) The 27 trials were divided into three blocks, each of nine trials. An analysis of variance with trial blocks as a repeated measures factor again showed no significant effects. (For trial blocks, $F (2,40) = 2.39, p>0.10$).

5.3.3 Discussion

i) Over all trials the percentage of attributes recalled correctly by the 30 seconds group was 31.8% and the percentage of slides correct was 61.7%. Melton's (1963b) study
showed 55% completely correct recall of CCC trigrams after 32 seconds (as read off his graph). Presentation conditions for his experiment and this pilot study were very similar; stimuli were presented by projector and read aloud at presentation, the slide to be remembered was presented for 1 second (Melton) or 0.8 seconds (pilot study), and the interval before the number slide was 0.7 seconds (Melton) or 1 second (pilot study). The slightly higher recall in this pilot study is presumably because order errors could not occur with size, colour, and pattern responses whereas they were possible with recall of CCC trigrams. Another reason could be the restricted number of attribute levels used in this study, as this raised the probability of a correct response by guessing.

Thus, the percentage recall found in this pilot study would appear to be within the range expected from Melton's experiment.

ii) The very slightly better recall for the 60 seconds group (84.2%) than for the 30 seconds group (81.8%) agrees with the results of Sheirer & Voss (1969). They found that recall for trigram pairs dropped to a minimum at about ten seconds and rose steadily but not steeply for longer retention intervals, regardless of the stimulus duration or the meaningfulness of CVC trigrams. This reminiscence effect occurred only when the retention interval was a between - S
Fig. 5  Comparison of the 30 \& 60 seconds groups.

\[ A = \text{Mean number of attributes correctly recalled per trial (maximum } = 3). \]
variable, as in this pilot study, and not a within-S variable. In most studies, an increase in retention interval results in an increase in the inter-trial interval. In a subsequent experiment in which these two intervals did not vary together, Sheirer & Voss found that it was the inter-trial interval which produced the phenomenon of reminiscence.

Figure 5 shows that in this pilot study PI did not build up more slowly for the 60 seconds group than for the 30 seconds group; it appeared to occur at the same rate but with less magnitude. Presumably the deleterious effects of increasing the retention interval and the advantageous effects of increasing the inter-trial interval balanced out.

Peterson & Gentile (1965) and Loess & Waugh (1967) have both shown improved performance with increased inter-trial intervals. Reminiscence in short-term memory has also been shown to occur with 5 item lists of paired associates (Peterson, 1966; Keppel & Underwood, 1967) but not with 2 item lists (Peterson & Peterson, 1962; Goggin, 1966).

5.3.4 Conclusion

In this pilot study a slight reminiscence effect was found. Therefore, increasing the retention interval would not make the task more difficult. Figure 5 shows very little change after the first nine trials. For these reasons it was decided to use a 30 seconds retention interval in the
second experiment and to give each S only nine trials instead of twenty seven.

5.4 Experiment 2

5.4.1 Design

As discussed in the introduction to this chapter, this experiment used a $2 \times 2 \times 2 \times 2 \times 2$ design with the following factors:

i) verbal/pictorial type of stimulus material
ii) compact/distributedness of stimuli
iii) voicing or silence at presentation
iv) sex
v) retention interval - 0 or 30 seconds.

There were 6 Ss in each cell of the above design giving a total of 192 Ss. However, each S occupied a unique cell in the total design because there were also three presentation orders and two sets of numbers for backwards counting. These two factors were not analysed. The three presentation orders were such that, summing over orders, all of the 27 stimulus slides were used and each attribute level was presented on each trial.

5.4.2 Stimuli

These were the same as in Experiment 1. Examples of each type are given in Appendix I.

5.4.3 Procedure

Each S was run individually. Ss were first given two
30 second periods of practice at backwards counting. These practice periods were slightly longer for some Ss who had difficulty with the task. Then there were three slide viewings for approximately 5 seconds each, and one at 0.8 seconds to familiarize Ss with the actual time of presentation. After inspection of a template showing the set of experimental stimuli, Ss were given 9 experimental trials. They attempted to count backwards to the 1 second beat of a metronome. As Wickens (personal communication, 1969) suggested, it would have been preferable to have recorded the amount of backwards counting on each trial. Murdock (1966) has shown that changes in performance on the interpolated activity may account for what is usually termed the growth of proactive inhibition and Talland (1967) has found inability to count backwards at the required rate to be related to failure to recall. However, the amount of manual operation required of the experimenter did not allow her to record the number of subtractions performed on each trial. Nor was there any a priori reason to expect differences in counting except perhaps between the 0 and 30 seconds groups. In the questionnaires, question 9 for the 0 seconds groups and questions 6, 7, and 17 for the 30 seconds groups related to backwards counting. A copy of the instructions is in Appendix V. After the experimental trials each subject filled in one of the questionnaires.
5.4.4 **Timing**

Timing for the 0 seconds and the 30 seconds groups was as follows:

<table>
<thead>
<tr>
<th>1 trial</th>
</tr>
</thead>
<tbody>
<tr>
<td>30 seconds</td>
</tr>
<tr>
<td>30 secs. group</td>
</tr>
<tr>
<td>30 secs. backwards counting</td>
</tr>
<tr>
<td>20 secs. group</td>
</tr>
<tr>
<td>20 secs. recall</td>
</tr>
<tr>
<td>30 seconds</td>
</tr>
</tbody>
</table>

↑ tone sounded for 1 second indicating that recall should begin

- 0.3 seconds presentation of a stimulus slide
- 1.0 seconds inter-slide interval
- 0.7 seconds presentation of a 3 digit number slide.

**5.4.5 Apparatus problems**

After about a fifth of the subjects had been run the Alphax mechanical shutter on the stimulus slide projector became quite unreliable. There was no gradual decrease in accuracy but on some trials there would be an apparently random failure so that the slide would be presented for 0.1 seconds or less instead of for 0.3 seconds. Results from several subjects had to be rejected because of such failures. Therefore, an electronic timing circuit was attached to the shutter which was left permanently on "brief". No more problems occurred with presentation time.
5.4.6 Questionnaire

In the questionnaire for the 30 seconds groups there were three main sections dealing with encoding at presentation, the interpolated activity, and recall. Only the section on encoding at presentation was given to the 0 seconds group. Data such as age and faculty were also collected. Numerical boundaries were suggested for the response categories, "Always", "Often", "Sometimes", "Never". Copies of the two questionnaires are in Appendix V.

5.5 Experimental Results

5.5.1 30 seconds group

Results from the 0 seconds and the 30 seconds groups could not be analysed together as the variances differed so widely. The range with 96 Ss was (0,3) for the 0 seconds group and (0,17) for the 30 seconds group.

As in Experiment 1, the relationship between slide errors and attribute errors was found to be highly significant. \( \chi^2 = 28.8, \, df = 3, \, p < 0.001 \) so only attribute errors were used in the analyses. Appendix III contains further details.

After the following results from analyses of variance the common types of errors, response orders, and labels for pictorial stimuli are briefly reported.

Table 13 gives the results of the analysis of variance on untransformed attribute errors. F ratios of less than
Table 13

Summary of Analysis of Variance for Expt. 2

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>P</th>
<th>VC</th>
<th>VR</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>297.5</td>
<td>29.23</td>
<td>&lt;.001</td>
<td>2.99</td>
<td>.215</td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>31.5</td>
<td>3.10</td>
<td>&gt;.05</td>
<td>0.22</td>
<td>.016</td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td>10.0</td>
<td>0.99</td>
<td>&gt;.25</td>
<td>-0.00</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>1</td>
<td>52.5</td>
<td>5.16</td>
<td>&lt;.05</td>
<td>0.44</td>
<td>.032</td>
</tr>
<tr>
<td>AB</td>
<td>1</td>
<td>2.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AC</td>
<td>1</td>
<td>15.8</td>
<td>1.55</td>
<td>&gt;.10</td>
<td>0.06</td>
<td>.004</td>
</tr>
<tr>
<td>AD</td>
<td>1</td>
<td>15.8</td>
<td>1.55</td>
<td>&gt;.10</td>
<td>0.06</td>
<td>.004</td>
</tr>
<tr>
<td>BC</td>
<td>1</td>
<td>3.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BD</td>
<td>1</td>
<td>0.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD</td>
<td>1</td>
<td>1.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABC</td>
<td>1</td>
<td>7.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABD</td>
<td>1</td>
<td>4.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACD</td>
<td>1</td>
<td>0.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BCD</td>
<td>1</td>
<td>2.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABCD</td>
<td>1</td>
<td>0.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>80</td>
<td>10.2</td>
<td></td>
<td></td>
<td>10.18</td>
<td>.731</td>
</tr>
</tbody>
</table>

A - Verbal/Pictorial Stimulus Type
B - Compact/Distributedness
C - Silence/Voicing at presentation
D - Sex
VC - Variance Component
VR - Variance ratio

One and negative variance components are omitted unless they occurred for main effects. There was a strong and highly significant verbal/pictorial stimulus effect (F (1,80) = 29.2, p<0.001, variance ratio = 21.5%), pictorial stimuli being recalled much better than verbal stimuli. Females recalled more than males (F (1,80) = 5.16, p<0.05) but this was a weak effect (variance ratio = 3.16%). The best transformation was at λ = 0.40 but this made little difference to the above
F ratios (30.6 and 5.13 respectively). Cell means for the untransformed data are shown in Table 14.

These results are contrary to those from Experiment 1 where only the compact/distributed factor was significant ($F(1,80) = 8.30, p<0.01$, variance ratio = 6.66%). The new factor of voicing at presentation was not significant ($F(1,80) = 0.99, p>0.25$, variance ratio = -0.00173%).

Table 14

Mean number of attribute errors

<table>
<thead>
<tr>
<th></th>
<th>Verbal Stimuli</th>
<th>Pictorial Stimuli</th>
</tr>
</thead>
<tbody>
<tr>
<td>Males</td>
<td>8.37</td>
<td>4.04</td>
</tr>
<tr>
<td>Females</td>
<td>6.09</td>
<td>3.37</td>
</tr>
<tr>
<td></td>
<td>7.23</td>
<td>3.71</td>
</tr>
</tbody>
</table>

Another analysis of variance was carried out with a fifth factor being the attribute errors decomposed into size, colour, and pattern scores. There was a significant attribute effect ($F(2,80) = 3.83, 0.01<p<0.05$) which was slightly increased at $\lambda = 0.20$ ($F = 4.01$).

Table 15

Mean number of errors

<table>
<thead>
<tr>
<th>Size</th>
<th>Colour</th>
<th>Pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>30 Secs. group</td>
<td>1.906</td>
<td>1.958</td>
</tr>
<tr>
<td>0 Secs. group</td>
<td>0.115</td>
<td>0.042</td>
</tr>
</tbody>
</table>
Fig. 6  Recall over trials.

$A =$ Mean number of attributes correctly recalled per trial
(maximum $= 3$).
ental trial. These practice viewings were necessary to show the pictorial distributed group what their unfamiliar stimuli would look like and to teach the pictorial compact group what the three sizes were. Judging by the 98.65% accuracy for the 0 seconds groups, the practice viewings served their purpose adequately.

A detailed analysis of errors revealed that 80.5% were confusion errors, 12.0% were blanks, and 7.5% were intrusions from the practice slides. Confusion errors were those where the response was incorrect but belonged to the set of attribute levels used in the experimental trials. A scattergram of blanks and confusion errors revealed no relationship between them whatsoever. It appears that guessing and leaving blanks represent two alternative strategies and it is unfortunate that the instructions did not emphasize that Ss must write down a size, colour, and pattern on each trial. Intrusion errors increased slightly as confusion errors increased. They also occurred almost exclusively in the first three trials; any occurring after that were repetitions of a previous intrusion error. The most common colour intrusion error was the only one to have been seen twice, "yellow", and the pattern intrusion errors were either the pattern seen twice, "wavy", or "mottled", which was similar to the experimental patterns, "dotted", "checked", and "striped".
While order of report for written responses does not necessarily reflect order of recall, there is presumably some relation between the two. It was found that for both verbal compact and pictorial distributed stimuli, the attributes were consistently reported from left to right - size, colour, then pattern. With verbal distributed stimuli there was a tendency to read the words from top to bottom rather than from left to right and this was even more marked for subjects who voiced the words at presentation. The opposite trend occurred with pictorial compact stimuli where Ss who were silent at presentation sometimes reported colour before size, whereas Ss who vocalized did not do this. All of these trends occurred for both the 0 and 30 seconds groups and results for these groups are combined in Table 16.

Table 16
Mean number of trials on which a certain response order was used - the maximum possible being nine

<table>
<thead>
<tr>
<th></th>
<th>Silent SCP</th>
<th>Silent CSP</th>
<th>Voiced SCP</th>
<th>Voiced CSP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal Compact</td>
<td>9.0</td>
<td>0.0</td>
<td>2.8</td>
<td>0.2</td>
</tr>
<tr>
<td>Verbal Distributed</td>
<td>6.7</td>
<td>2.3</td>
<td>4.0</td>
<td>5.0</td>
</tr>
<tr>
<td>Pictorial Compact</td>
<td>7.2</td>
<td>1.8</td>
<td>3.9</td>
<td>0.1</td>
</tr>
<tr>
<td>Pictorial Distrib.</td>
<td>9.0</td>
<td>0.0</td>
<td>9.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

SCP - Ss responded size, colour, then pattern
CSP - Ss responded first with colour

This analysis of order of report does not explain why pattern
was the attribute best recalled but it does suggest a need for caution in interpreting differences between recall for size and for colour.

Ss who saw pictorial stimuli were not told what words to use to describe these stimuli. Hence an analysis was carried out to see what words were used. Tables listing these words are given in Appendix VI. They show that Ss who saw words and Ss who saw pictures used much the same words for sizes and colours but differed on patterns. It is unlikely that this affected memory for these attributes as there was no interaction between verbal/pictorial stimuli and attributes for the 30 seconds or the 0 seconds groups.

5.5.2 0 seconds group

The range of scores for 96 Ss was (0,3) and the average percentage of attribute errors was 1.35%. The number of attribute errors declined over the nine trials.

Table 17

Number of attribute errors made on each trial by 96 Ss

<table>
<thead>
<tr>
<th>Trials</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Errors</td>
<td>9</td>
<td>7</td>
<td>3</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

A, the number of attribute errors for each subject, may have been distributed binomially but not as the sum of 27 random binomial variables, all with probability p.
Table 18

Mean number of errors for each stimulus condition

<table>
<thead>
<tr>
<th>Verbal (V)</th>
<th>Pictorial (P)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compact (C)</td>
<td>Distributed (D)</td>
</tr>
<tr>
<td>0.125</td>
<td>0.500</td>
</tr>
</tbody>
</table>

$\chi^2$ tests on the number of errors per subject revealed no differences between these groups or between groups who vocalized and groups who did not vocalize at presentation.

Multinomial tests for $V_C$, $V_D$, $P_C$, and $P_D$ groups showed no differences between attributes for $V_C$ ($p = 0.338$) more pattern errors for $V_D$ ($p = 0.000$) and $P_D$ ($p = 0.000$), and more size errors for $P_C$ ($p = 0.047$).

Results from the 0 seconds group showed perceptual errors to be rare.

5.6 Questionnaire Results

Copies of the questionnaires for the 0 and 30 seconds groups are in Appendix VI. The figures given underneath each question are the percentages of the 96 Ss who responded with each category.

Results for the 0 seconds groups are presented and compared with those for the 30 seconds groups. Then questionnaire results from Experiments 1 and 2 are compared. Ss' estimates of their percent correct recall are shown in relation to their actual scores. Other results for the
30 seconds groups are discussed and the results of a factor analysis on questionnaire responses and error scores are presented. This factor analysis led to a more detailed investigation of encoding, reported in section 5.6.7.

5.6.1 0 seconds group

Results from all questions were cross tabulated with each other, the experimental factors, and attribute error scores. Apart from relationships within the census data (e.g. age and year at University) only questions 1 & 2, and 1 & 3 were related. The more Ss "thought" in words (Q1) the more likely they were to "hear" the words (Q2) ($\chi^2 = 22.4, \text{df} = 2, p<0.001$). Also, the more Ss "thought" in words the less likely they were to redundantly encode the shape of what they saw (Q3) ($\chi^2 = 4.45, \text{df} = 1, 0.02<p<0.05$). The verbal/pictorial and compact/distributed factors did not affect questionnaire answers and the voicing at presentation factor affected only the number of times the slide was repeated before being written down (Q4). Ss who were silent at presentation repeated what was on the slide more often than those who vocalized ($\chi^2 = 8.5, \text{df} = 1, 0.001<p<0.01$). This voicing factor did not affect "thinking" in words (Q1) or "hearing" words (Q2), probably because most Ss did these things anyway. However there is the possibility that Ss who had to vocalize could not understand these
questions. Sex was related only to some of the census data such as faculty (e.g. Arts or Science).

None of the questions related to the number of attributes errors; as there were so few errors this result is not unexpected.

5.6.2 Comparison of 0 and 30 seconds groups

For both groups Q5, on the number of rehearsals, was slightly ambiguous: an answer of "3" could mean three verbal repetitions at presentation for each trial or else one repetition on only three of the nine trials with no repetitions on the other six trials. Therefore, only Q4, a "Yes/No" question on rehearsal, could be used in the analysis.

Comparison $\chi^2$'s were carried out on the census data and on questions 1, 2, 3, 4, 15, 16, & 17 for the 30 seconds group and questions 1, 2, 3, 4, 7, 8, & 9 for the 0 seconds group.

The 30 seconds group "heard" words (Q2) significantly less than did the 0 seconds group ($\chi^2 = 9.46, \text{df} = 3, 0.02 < p < 0.05$) and there was a slight but nonsignificant trend in the same direction for "thinking" of the stimulus in words (Q1).

The 30 seconds group also repeated what was on the slide significantly less than did the 0 seconds group ($\chi^2 = 4.4, \text{df} = 1, 0.02 < p < 0.05$).
Not surprisingly, the 30 seconds group reported that they "muddled the slide they had just seen with the previous one" much oftener than did the 0 seconds group \( (\chi^2 = 31.2, \ df = 2, \ p < 0.001) \). This corresponds to the different error rates for the two groups - 20.2% for the 30 seconds group and 1.35% for the 0 seconds group.

In the open comments question, 11 of the 13 who reported forming an image from the verbal stimuli came from the 30 seconds group (using a binomial test, \( p = 0.011 \)). While open questions do not provide unequivocal evidence, it does seem likely that images are less likely to be formed for immediate than for delayed verbal recall. Further evidence for the 30 seconds group comes from Q11, "Did you see an image of a picture (at recall)?". Of the 48 Ss who saw verbal stimuli, 23 replied "Never", 12 replied "Sometimes", 6 replied "Often", and 2 replied "Always".

Overall, it appears that there is slightly less verbal encoding and more pictorial encoding for verbal material when material has to be remembered for some time. This is not surprising as the 0 seconds task was just a copying task for the verbal stimulus groups and an immediate translation to words for the pictorial stimulus groups.

5.6.3 Comparison with Experiment 1

In Experiment 1 there was only one question for each type of encoding; in Experiment 2 there were questions on
encoding at presentation and on the form of recall. Also, in Experiment 2, half of the Ss vocalized at presentation. Although there was no difference between silent and vocalizing Ss on any of the questions discussed in this section it was thought advisable to compare only the results from silent Ss with those from Experiment 1 for the questions relating to encoding at presentation. This was because it was suspected that Ss who had to vocalize were confused by the questions on verbal and auditory encoding at presentation.

Table 19
A comparison of encoding in Experiments 1 and 2

<table>
<thead>
<tr>
<th>Expt. 1</th>
<th>Expt. 2</th>
<th>Relation</th>
<th>$\chi^2$</th>
<th>df</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q2</td>
<td>Q1</td>
<td>(1) &gt; (2)</td>
<td>9.9</td>
<td>2</td>
<td>p&lt;0.01</td>
</tr>
<tr>
<td>&quot;thinking&quot; in words (at presentation)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q3</td>
<td>Q2</td>
<td>(1) &gt; (2)</td>
<td>11.3</td>
<td>3</td>
<td>p=0.01</td>
</tr>
<tr>
<td>&quot;hearing&quot; words (at presentation)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q2</td>
<td>Q8</td>
<td>(1) &gt; (2)</td>
<td>26.4</td>
<td>3</td>
<td>p&lt;0.001</td>
</tr>
<tr>
<td>&quot;thinking&quot; in words (at recall)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q3</td>
<td>Q9</td>
<td>(1) &gt; (2)</td>
<td>40.4</td>
<td>3</td>
<td>p&lt;0.001</td>
</tr>
<tr>
<td>&quot;hearing&quot; words (at recall)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q5</td>
<td>Q11</td>
<td>(1) = (2)</td>
<td>6.2</td>
<td>3</td>
<td>p=0.10</td>
</tr>
<tr>
<td>&quot;seeing&quot; (pictures)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q5</td>
<td>Q10 &amp; Q11</td>
<td>(1) &lt; (2)</td>
<td>14.8</td>
<td>3</td>
<td>p&lt;0.01</td>
</tr>
<tr>
<td>&quot;seeing&quot; (pictures or words)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
It would seem that there was less verbal and auditory encoding in Experiment 2 and more visual encoding. Results for Q10 and Q11 were independent ($x^2 = 6.57, \text{df} = 4, p > 0.10$) and so an imagery score was formed by summing over both questions. Only two Ss received a summed score of greater than 3 (0 = never, 1 = sometimes, 2 = often, and 3 = always; this was the scoring system used for each question). These two Ss were incorporated in the "3" group and the comparison with Experiment 1 was carried out as for single question.

In conclusion then, just as the two experiments gave different results in terms of error measures, so did they differ in the encoding used by Ss.

5.6.4 Expected Percentage Correct

Questions 13 and 14 for the 30 seconds group required Ss to estimate the percentage of sizes, colours, patterns, and total slides that they recalled correctly. These estimates were regressed on the actual error scores using the IBM Scientific Subroutine Sample Program, POLREG. Only linear components were significant for all four regressions. In the following equations the number of errors has been converted to the percentage correct to make interpretation easier.
\[ X = \text{percentage correct} \]

\[ \text{EX} = S\text{'s estimate of } X \]

<table>
<thead>
<tr>
<th>Attribute</th>
<th>EX equation</th>
<th>F value</th>
<th>df</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size</td>
<td>(0.77 X - 0.93)</td>
<td>31.3</td>
<td>1,94</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Colour</td>
<td>(0.67 X + 6.9)</td>
<td>26.4</td>
<td>1,94</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Pattern</td>
<td>(0.77 X - 9.9)</td>
<td>28.1</td>
<td>1,94</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Slides</td>
<td>(0.66 X + 8.8)</td>
<td>48.5</td>
<td>1,94</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Thus Ss' estimates of percentages correct were closely related to their actual percentages but this relationship was not EX = X.

5.6.5 30 seconds group

The results discussed in this section are only those not covered subsequently in the factor analysis (5.6.6) or the discussion of encoding (5.6.7).

There was some suspicion that Ss who had to vocalize at presentation were perplexed by questions 1 and 2 on verbal and auditory encoding. One would expect such Ss to have replied that they always thought of the stimulus in words and that they always "heard" these words. This was not the case. Their responses were not significantly different from those of the silent Ss \((\chi^2 = 5.30, df = 3, p>0.10; \chi^2 = 6.02, df = 3, p>0.10)\).

<table>
<thead>
<tr>
<th>Table 20</th>
</tr>
</thead>
<tbody>
<tr>
<td>Responses from Ss who vocalized at presentation</td>
</tr>
<tr>
<td>Never</td>
</tr>
<tr>
<td>Q1 - &quot;thinking&quot; in words</td>
</tr>
<tr>
<td>Q2 - &quot;hearing&quot; words</td>
</tr>
</tbody>
</table>
The slight bimodality shown in Table 20 may suggest that Ss decided that vocalizing either was or was not what was meant by these questions. However the same trends show up in the recall in words (Q9) and "hearing" at recall (Q9) questions. There is no reason why Ss who vocalized should misunderstand these questions, although there is always the possibility that they felt that they had to answer both presentation and recall questions in the same way to be consistent. One S spontaneously commented that in spite of having to vocalize at presentation he did not really think of the stimuli in words. This would seem to mean that he placed priorities on visual rather than verbal encoding. More information of this kind is needed to show the ways in which questions 1 and 2 were interpreted by Ss who had to vocalize at presentation.

It is interesting to note that not all Ss interpreted Q4 to mean verbal repetition. Four of the thirteen Ss who "never" thought in words replied that they did manage to repeat what was on the stimulus slide before that number slide appeared.

Responses to question 6 and 7 on rehearsal during backwards counting are of interest as an estimate of how successfully this interpolated task prevents rehearsal. Measures of the number of subtractions have been found to relate to differences in recall between groups (Nickens & Simpson, 1968)
but the only studies in which Ss were asked about covert rehearsal were those of Groninger (1966) and Talland (1967). Groninger found that the amount of reported rehearsal while counting significantly affected recall and that over four trials the percentage of Ss reporting covert rehearsal dropped from 60% to 40%. One would expect that as Ss practiced backwards counting they would improve and therefore find it easier to carry out the task and rehearse. On this argument, the percentage reporting covert rehearsal would be expected to increase over trials. It is possible that when Groninger's Ss were asked about covert rehearsal after every trial they gained the impression that they should not have rehearsed. Hence, some of them may either have tried not to rehearse while counting or just not reported what they did. Talland merely noted that few Ss in his experiment said that they tried to rehearse while counting or reading.

Results for Experiment 2 are given in Table 21.

**Table 21**

<table>
<thead>
<tr>
<th>Percentage of Ss responding with each category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Never</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>Q6 - repeating while counting</td>
</tr>
<tr>
<td>Q7 - visualizing while counting</td>
</tr>
</tbody>
</table>
Responses to question 6 and 7 were independent ($\chi^2 = 0.003$, df = 1, $0.90 < p < 0.95$). However, only 40 of 96 Ss reported that they never repeated or visualized the stimulus to themselves while counting backwards. Thus any interpretation of errors which assumes that these indicate a pure measure of decay over time is incorrect, quite apart from the interference arguments.

The type of rehearsal during backwards counting was dependant on the type of stimulus material. Verbal stimulus Ss nearly repeated more often than Pictorial stimulus Ss ($\chi^2 = 3.59$, df = 1, $\chi^2_{\text{crit.}} = 3.89$ for $\alpha = 0.05$) and the reverse was true for visualizing ($\chi^2 = 6.65$, df = 1, $p = 0.01$).

"Seeing" words at recall (Q10) might be expected to be rather unlikely for Ss who had seen pictorial stimuli. It was not common but it did occur, although significantly less than for Ss who saw verbal stimuli ($\chi^2 = 11.6$, df = 2, $0.001 < p < 0.01$).

**Table 22**

Response frequencies for Q10

<table>
<thead>
<tr>
<th></th>
<th>Never</th>
<th>Sometimes</th>
<th>Often</th>
<th>Always</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pictorial Stimuli</td>
<td>36</td>
<td>10</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Verbal Stimuli</td>
<td>21</td>
<td>16</td>
<td>9</td>
<td>2</td>
</tr>
</tbody>
</table>
Storage of the visual characteristics of stimuli is evidenced by responses to "seeing" words (Q10) for Verbal stimulus Ss and "seeing" pictures (Q11) for Pictorial stimulus Ss. A comparison indicates that the visual characteristics of words are remembered less often than the visual characteristics of pictures ($\chi^2 = 18.67, df = 3, p<0.001$). This is further evidence for the view that the visual characteristics of words are used mainly to process meaning and possibly to generate sound or vocal kinaesthesis; this notion implies that the visual characteristics themselves are of little importance and are seldom stored.

One weakness of this questionnaire was that Verbal stimulus Ss were not asked if they formed an image from the words at presentation. Nor was there a direct question on the type of image, if any. The only information on this came from the open comments question (Q16). Some Ss formed an image of a referent object like a large red dotted handkerchief while others formed a more geometrical type of image just showing the appropriate attribute levels.

Table 23
Types of images formed from verbal stimuli (Q16)

<table>
<thead>
<tr>
<th>Referent Object</th>
<th>Abstract Referent</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 seconds group</td>
<td>0 (1)</td>
</tr>
<tr>
<td>30 seconds group</td>
<td>5</td>
</tr>
</tbody>
</table>

(1) - number of Ss.
As twenty of the forty eight 30 seconds Ss who saw verbal stimuli responded that they at least sometimes "saw" a picture at recall, and two of the eleven who mentioned images replied "never" to this question, there must be a minimum of eleven Ss for whom the type of images are not known. Not all attributes were visualized - one S said that she visualized only colour and another visualized only colour and pattern.

Referent objects mentioned by Ss were "blue dotted handkerchief", "large red checked tablecloth", "blue striped pyjamas", "large green striped candy stick", "flag", and "cloudy sky". The most unified system of imagery was started after one or two trials by a boy who encoded size, colour, and pattern on bikinis. He had less errors than the other five males in his experimental condition!

Because of the small number of Ss who discussed imagery in the open question it was not worthwhile comparing recall scores for those who did and those who did not discuss imagery. Q11, on "seeing" pictures, serves as a more reliable indication of the effectiveness of imagery.

One subject was rejected because he was found to have eidetic imagery so that he continued to "see" the pictorial stimulus until he had to recall. Accordingly, for him, recall was just verbal description. He made no errors. His phenomenal ability in this respect was investigated
in subsequent sessions using art material. A report on results from these sessions is being prepared. This subject was a Fijian student; all other Ss were native speakers of English and had been brought up in a Western culture.

5.6.6 Factor Analysis

This was used, as in Experiment 1, to help sort out relations between experimental variables, error measures, and questionnaire replies. Only the data for the 30 seconds group was analysed in this way. All questions were included except for questions 5, 12, 16, and 18. Q5 was the ambiguous question on rehearsal, Q12 was concerned with attribute importance, and Q16 and Q18 were open questions. One error measure, A, the number of attribute errors, was included; size, colour, and pattern errors summed to A and so could not be included as well as A. All census data was used. There were 25 variables and 96 Ss.

Eight factors with eigenvalues of greater than one were extracted and rotated. The eigenvalues, percentage eigenvalues, and all rotated factor loadings greater than 0.30 are listed in Table 24. As in Experiment 1, there is no general factor and the variance taken up by each factor flattens out after the first two factors, being only 7% for factors III, IV, and V and 5% for factors VI, VII, and VIII.

Factor I appears to be a general recall factor, slightly more related to the expected percentage of correctly
recalled slides than to the number of attributes recalled. Recall is higher for pictorial stimuli, as found in the analysis of variance, and is related to hearing words at recall and to seeing pictures at recall. It is inversely related to improvement in counting. As the regression analyses (5.6.4) showed, expected scores are closely related to actual scores.

Factor II looks like a general encoding factor. All the verbal and auditory encoding questions have high negative loadings on it and seeing pictures at recall has a positive loading. The negative loading for Q17 suggests that the more visual encoding takes place, the less people feel that they improve at backwards counting. The different signs of the loadings for verbal and pictorial recall and encoding indicate that these are alternative strategies. As shown by factor I, the more either sort of encoding is used, the better the recall.

Factor III would seem to be concerned with vocalizing at presentation. It is inversely related to rehearsal at presentation. This is understandable in the light of Landauer's (1962) finding that subvocal repetition is faster than vocal repetition except for much practised material. Question 10, on seeing words at recall, has a negative loading on this factor. Presumably one is less likely to retain the visual characteristics of verbal stimuli when
Table 24
Rotated Factor Loadings for Expt. 2.

<table>
<thead>
<tr>
<th>Variables</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>V</th>
<th>VI</th>
<th>VII</th>
<th>VIII</th>
<th>h²</th>
</tr>
</thead>
<tbody>
<tr>
<td>V/P</td>
<td>49</td>
<td></td>
<td></td>
<td>39</td>
<td>37</td>
<td>-35</td>
<td>69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C/D</td>
<td></td>
<td>63</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>54</td>
</tr>
<tr>
<td>S/A</td>
<td>82</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>72</td>
</tr>
<tr>
<td>M/F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>78</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td>-30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>77</td>
</tr>
<tr>
<td>Year</td>
<td></td>
<td>-71</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>55</td>
</tr>
<tr>
<td>Faculty</td>
<td></td>
<td>35</td>
<td></td>
<td>-33</td>
<td>-54</td>
<td></td>
<td></td>
<td></td>
<td>69</td>
</tr>
<tr>
<td>F/Part</td>
<td></td>
<td></td>
<td></td>
<td>-68</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>53</td>
</tr>
<tr>
<td>Q1</td>
<td>-84</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>74</td>
</tr>
<tr>
<td>Q2</td>
<td>-51</td>
<td></td>
<td></td>
<td>49</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>66</td>
</tr>
<tr>
<td>Q3</td>
<td></td>
<td></td>
<td>-75</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>71</td>
</tr>
<tr>
<td>Q4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>66</td>
</tr>
<tr>
<td>Q5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>71</td>
</tr>
<tr>
<td>Q6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-36</td>
<td>53</td>
<td></td>
<td>56</td>
</tr>
<tr>
<td>Q7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>72</td>
<td></td>
<td></td>
<td>57</td>
</tr>
<tr>
<td>Q8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-77</td>
<td></td>
<td></td>
<td>65</td>
</tr>
<tr>
<td>Q9</td>
<td>34</td>
<td>-63</td>
<td></td>
<td>-36</td>
<td>-51</td>
<td>-31</td>
<td></td>
<td></td>
<td>73</td>
</tr>
<tr>
<td>Q10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>73</td>
</tr>
<tr>
<td>Q11</td>
<td>38</td>
<td>53</td>
<td></td>
<td></td>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td>71</td>
</tr>
<tr>
<td>Q12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>71</td>
</tr>
<tr>
<td>Q13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>71</td>
</tr>
<tr>
<td>Q14</td>
<td>91</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>84</td>
</tr>
</tbody>
</table>

Eigenvalues  | 4.85 | 2.78 | 1.93 | 1.80 | 1.53 | 1.42 | 1.32 | 1.16 |

% e.v.        | 19   | 12   | 07   | 07   | 07   | 05   | 05   | 05   |

Decimal points have been omitted except for the eigenvalues.

V/P = verbal/pictorial factor  C/D = compact/distributedness
S/A = silent/aloud voicing    M/F = male/female factor
F/Part = full/part-time dichotomy  A = attribute errors
% e.v. = percentage eigenvalues  h² = communalities
one is forced to make an auditory transformation.

Factor IV looks like an age factor and is theoretically irrelevant. The pattern of loadings for Factor V is difficult to understand. Probably it is not worth trying to interpret this factor, let alone factors VI, VII, and VIII. They all take up very little variance and have eigenvalues close to 1 (c.f. Horn, 1966).

To conclude, the factor analysis has produced two meaningful but small factors and one meaningful but very small one. The pictorial nature of the stimuli, the amount of verbal or pictorial encoding, and the reported lack of improvement at the interpolated activity, all relate highly to the recall factor. The second factor shows verbal and pictorial encoding to be alternative ways of storing information for later recall. Finally, vocalizing is inversely related to amount of rehearsal at presentation and the likelihood of seeing words at recall.

5.6.7 A Comparison of Encoding for Verbal and Pictorial Stimuli

A more detailed analysis of encoding for verbal and pictorial stimuli was carried out because Factor 2 suggested that verbal/auditory encoding and visual encoding were alternative forms. This would appear to contradict the hypothesis that pictorial stimuli are remembered better because they are encoded both verbally and visually. However, Factor 2 need not imply that alternatives are mutually
exclusive. It seems more reasonable to interpret it as a continuum with purely verbal/auditory encoding at one end, purely visual encoding at the other end, and mixed strategies in the middle (c.f. Richardson, 1969, Chapter 4).

Only the questions which had loadings greater than 0.30 on Factor 2 were considered in this analysis.

**Table 25**

A comparison of encoding for verbal and pictorial stimuli

<table>
<thead>
<tr>
<th>Question</th>
<th>Content</th>
<th>Relation</th>
<th>$\chi^2$</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>&quot;thinking&quot; in words at presentation</td>
<td>$V = P$</td>
<td>6.18</td>
<td>3</td>
<td>$p&gt;0.10$</td>
</tr>
<tr>
<td>2</td>
<td>&quot;hearing&quot; words at presentation</td>
<td>$V = P$</td>
<td>2.41</td>
<td>3</td>
<td>$p&gt;0.30$</td>
</tr>
<tr>
<td>8</td>
<td>&quot;thinking&quot; in words at recall</td>
<td>$V = P$</td>
<td>1.36</td>
<td>3</td>
<td>$p&gt;0.70$</td>
</tr>
<tr>
<td>9</td>
<td>&quot;hearing&quot; words at recall</td>
<td>$V = P$</td>
<td>1.82</td>
<td>2</td>
<td>$p&gt;0.30$</td>
</tr>
<tr>
<td>11</td>
<td>&quot;seeing&quot; pictures at recall</td>
<td>$V &lt; P$</td>
<td>18.67</td>
<td>3</td>
<td>$p&lt;0.001$</td>
</tr>
</tbody>
</table>

As Table 25 shows, Pictorial Ss verbally encoded as much as Verbal Ss but they visually encoded in pictorial form much more than Verbal Ss. Visual encoding of words did not seem to relate to recall and so was ignored here. Summing scores for questions 8 and 11 and for questions 9 and 11 showed that Pictorial Ss had significantly higher levels of encoding ($\chi^2 = 15.65$, df = 3, $0.001<p<0.01$ and
\( \chi^2 = 28.7, \text{ df } = 4, \text{ p}<0.001 \) respectively.

These results lend some support to the hypothesis that pictorial stimuli are recalled better than verbal stimuli because they are encoded twice (visually and verbally) whereas verbal stimuli are encoded only once (verbally). The fault with this notion is that it oversimplifies; the important factor is not single versus dual encoding but what Paivio & Csapo (1969) call the "summative availability" of both visual and verbal encoding. In this experiment the summative availability was higher for pictorial stimuli.

All of the verbal/auditory encoding questions were closely related and all were inversely related to recall in pictures, as Factor 2 revealed.

Thus, it appears that there is an inverse relationship between the two sorts of encoding but that pictorial stimuli are more likely to be encoded both ways than verbal stimuli. Presumably naming is a more common activity than forming images. This agrees with the work on latencies for naming and image arousal (Fraisse, 1964; Paivio, 1966) discussed in Paivio & Csapo (1969).

5.7 Discussion

In this Experiment there was a very significant verbal/pictorial effect. Subjects also showed some insight into how well they recalled. The more pictorial or verbal encoding reported, the better the recall. These two types of encoding appeared to be alternatives but pictorial stimuli were encoded
both ways more than were verbal stimuli. This greater amount of encoding was put forward as the explanation of the superior recall for pictorial stimuli. Results for the 0 seconds group showed slightly more errors for pictorial stimuli; hence the difference in recall errors is not due to perceptual errors.

Compact/distributedness had no effect on recall scores \( (F = 3.10 \text{ df } = 1,80 \ 0.05 < p < 0.10) \) or on perceptual error scores. This contrasts with Experiment 1 where recognition scores were lower for distributed stimuli. The difference between the two experiments may be a function of the types of tasks used; alternatively, it may indicate that the lower scores for distributed stimuli in Experiment 1 were due to the response templates being compact and not to memory for the stimuli themselves.

Voicing at presentation had no effect on recall scores \( (F = 0.99 \text{ df } = 1,80 \ p > 0.25) \). However, it did appear as Factor III in the factor analysis where it was inversely related to repetition of the stimulus during presentation. The negative loading on Q10 indicates that the visual characteristics of words are less likely to be retained when the words are voiced at presentation.

Females were significantly better than males at the memory task \( (F = 5.16 \text{ df } = 1,80 \ 0.01 < p < 0.05) \). It is not understood why this was so.
There was a significant attribute effect which was contrary to the pattern of perceptual errors. An analysis of order of report showed a need for caution in interpreting differences in recall for colour and pattern. Detailed analyses of errors and response words cast no light on any other effects.

Recall declined and then rose again over trials suggesting the combined effects of PI and practice.

Results from the questionnaire indicate that all forms of encoding set out in the diagrammatic model for visual input were used. Some forms were much less common than others. There is one major problem in using evidence from the questionnaire for the model; many of the questions relate to imagery at recall and not to encoding at presentation. The very strong relationships between the verbal and auditory questions on recall and presentation indicate that it is reasonable to infer encoding from recall imagery. Nonetheless it is unfortunate that there was no question on imagery at presentation. For instance, those Ss who saw pictorial stimuli and reported "seeing" words at recall may not have stored images of words — they may have recalled auditorily and formed an image from such recalled information. Replies to questions on encoding at presentation can, however, be difficult to interpret. For example, if a subject saw words and yet replied that he only "sometimes" thought in words,
what does this mean? Is his reply related to the importance rather than to the frequency of that form of encoding?

A clearer picture of the relative likelihoods of the possible forms of encoding is obtained from the rank ordering given below. Q8 was used as the measure of verbal encoding and Q9 as the measure of auditory encoding. Q10 related to seeing words and Q11 to seeing pictures. The term "encoding" is used although "imagery at recall" might be preferable.

**Rank Orders of forms of encoding**

<table>
<thead>
<tr>
<th>Verbal Stimuli</th>
<th>Pictorial Stimuli</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Verbal encoding</td>
<td>1. Visual encoding of pictures</td>
</tr>
<tr>
<td>2. Auditory encoding</td>
<td>2. Verbal encoding</td>
</tr>
<tr>
<td>3. Visual encoding of words</td>
<td>3. Auditory encoding</td>
</tr>
<tr>
<td>4. Pictorial encoding of referents</td>
<td>4. Visual encoding of words</td>
</tr>
</tbody>
</table>

Table 26 shows these orders in more detail.

**Table 26**

**Response frequencies**

<table>
<thead>
<tr>
<th>Verbal Stimuli</th>
<th>Never</th>
<th>Sometimes</th>
<th>Often</th>
<th>Always</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Q8</strong></td>
<td>7</td>
<td>10</td>
<td>12</td>
<td>19</td>
</tr>
<tr>
<td><strong>Q9</strong></td>
<td>21</td>
<td>16</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td><strong>Q10</strong></td>
<td>21</td>
<td>16</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td><strong>Q11</strong></td>
<td>28</td>
<td>12</td>
<td>6</td>
<td>2</td>
</tr>
</tbody>
</table>
Pictorial Stimuli

<table>
<thead>
<tr>
<th>Question</th>
<th>Never</th>
<th>Sometimes</th>
<th>Often</th>
<th>Always</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q8</td>
<td>11</td>
<td>10</td>
<td>12</td>
<td>15</td>
</tr>
<tr>
<td>Q9</td>
<td>18</td>
<td>13</td>
<td>14</td>
<td>3</td>
</tr>
<tr>
<td>Q10</td>
<td>36</td>
<td>10</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Q11</td>
<td>5</td>
<td>17</td>
<td>13</td>
<td>13</td>
</tr>
</tbody>
</table>

5.8 What is verbal encoding?

So far, verbal encoding has been used rather loosely, sometimes referring to "thinking" in words and sometimes acting as an overall term for both "thinking" and "hearing". In Experiments 1 and 2 Ss have indicated:

i) less "hearing" than "thinking" in words

ii) the more "thinking" reported, the more "hearing" reported.

Most of the studies comparing pictures and words have just talked of visual and verbal encoding without further specification (e.g. Jenkins, Neale, & Deno, 1967; Paivio & Casp, 1969).

In section 1.5 the auditory, articulatory, and abstract systems for short-term verbal memory were reviewed. As Wicklagren (1969b) points out, there is as yet no clear way of distinguishing between these three possible systems. In Experiments 1 and 2 it was not expected that Ss would be familiar with these possibilities and so questions on each would have to have been couched in common terms such as
"hearing", "saying", and "knowing without hearing or saying". It would have been interesting to see how Ss coped with such questions although their responses could not settle the issue. Instead it was assumed that "thinking" in words covered all three possibilities and that "hearing" included both auditory and articulatory possibilities.

5.9 **Summary**

In Experiment 2, in which Ss had to recall one slide after 30 seconds of backwards counting, pictorial stimuli were recalled much better than verbal stimuli. This appeared to be because, while both types of stimuli were verbally encoded to the same extent, pictorial stimuli were more often also encoded visually.
6.1 Comparison of Experiments 1 & 2

The theoretically important results from Experiments 1 and 2 are those concerned with the verbal/pictorial stimulus and response factors and the reported forms of encoding. As the compact/distributedness of stimuli did not affect encoding it is of no further interest.

There are some problems in a comparison of Experiment 1 and Experiment 2. The type of response required in each experiment was different. It is not clear how much Experiment 1 involved recall and how much it involved recognition but it was certainly not the same task as in Experiment 2 where only recall was required. Norman (1968) and Shiffrin & Atkinson (1969) suggest that recall differs from recognition in the amount of searching through storage, recall usually involving a search through a larger array. In both experiments Ss were all familiar with the total set of possibly correct answers; evidence for this in Experiment 2 comes from the low level of practice intrusions and the complete absence of other intrusions. Therefore the array to be searched was the same size for both experiments. Also, in Experiment 1 the distractor items were such that partial recognition could not pass for total recognition as
may happen with nonsense syllables (e.g. XOZ with distractors XIK and ZEL; c.f. McNulty, 1965). Accordingly, even if Experiment 1 involved purely recognition responses there is no reason to believe these would be very different from recall responses, let alone to think that recognition and recall would be differentially affected by the type of stimulus. Experimental confirmation would be desirable but in the absence of such confirmation it will be assumed that differences between Experiment 1 and Experiment 2 are not a function of the type of response required. The absence of a response template type effect in Experiment 1 also aids comparison of the two experiments as well as being intrinsically interesting.

There were small differences in presentation time for stimuli. Both presentation time and the inter-slide interval were longer in the second experiment and yet the rate of perceptual errors was higher (1.35% compared with 0.86%). This was probably because the task in Experiment 1 was self-paced.

It seems unlikely that the theoretically important results from each experiment were caused by the above mentioned differences between them.

The prime difference would seem to be the list versus single item distinction. In Experiment 1 three, three-attribute items were presented in sequence; in Experiment 2
only one such item was presented on any one trial. Admittedly, perception of a three-attribute item may be sequential but the attributes are all unique whereas a sequence of levels on one attribute corresponds to the usual notion of a list. The boundary zone between a complex item and a list is indeed unclear but Experiments 1 and 2 would seem to be on opposite sides of this zone. In the subsequent discussion Experiment 1 will be said to involve sequential presentation of a list whereas Experiment 2 will be referred to as an example of one item presentation.

6.2 Comparison with other studies and a theoretical discussion

Not only are there differences between Experiments 1 and 2 but there are even wider differences between many of the other studies comparing verbal and pictorial stimuli. However, it does not seem to distort the complex methodological interrelations too much to place groups of studies in the cells of the following table.

\[ P = \text{pictorial stimuli} \]
\[ V = \text{verbal stimuli} \]

\( P > V \) means that pictorial stimuli were "remembered" better than verbal stimuli, as measured by recognition or recall.
**Table 27**

<table>
<thead>
<tr>
<th>Sequential Presentation &amp; Serial Recall</th>
<th>Sequential Presentation &amp; Free Recall or Recognition*</th>
<th>Presentation &amp; Recall of only one item</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of attributes not equated for</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P &lt; V at fast rates, otherwise</td>
<td>P = V at fast rates, otherwise</td>
<td>P &gt; V (?) Dominowski &amp; Gadlin (1968)</td>
</tr>
<tr>
<td>Paivio &amp; Caspo (1969)</td>
<td>P &gt; V</td>
<td></td>
</tr>
<tr>
<td>Number of Attributes equated for</td>
<td>P = V Experiment 1</td>
<td>P &gt; V Experiment 2</td>
</tr>
<tr>
<td>--------------------------------------</td>
<td>------------------------------------------------------</td>
<td>---------------------------------------</td>
</tr>
</tbody>
</table>

* Lists of single items or lists of paired-associates

Some of the exceptions to the general trend for superior memory with pictorial stimuli, which were discussed in Chapter 3, have been omitted from this table as they were not related to the serial nature of the tasks involved. The question mark beside Dominowski & Gadlin's result is to indicate that the relationship reported is contrary to their interpretation (3.2). The grouping together of paired-associate, free recall, and recognition studies is debatable but not unreasonable for the purposes of this discussion - items in all three types of studies are presented serially, in lists.

In Experiment 1 the response order was free but the serial position of each item had to be reported. For serial recall items have to be reported in order of presentation;
for free recall order is irrelevant. Thus Experiment 1 was intermediate between these two methods although it is grouped with free recall in Table 27.

This tabular presentation of studies suggests that:

i) the more serial the task, the less superior is memory for pictorial stimuli,

ii) when the number of attributes carried by verbal and pictorial stimuli is equated, the superiority of pictorial stimuli is diminished, and

iii) the superiority of pictorial stimuli is also affected by the rate of presentation, the crucial rate being between three and five items per second (Paivio & Csapo, 1969).

These conclusions must remain somewhat tentative because of the empty cell in Table 27 and the wide variety of methods only partially represented in this table. Nonetheless they fit well within a theoretical framework in which vision is assumed to be primarily a parallel processing system whereas the verbal complex of audition and speech is assumed to be specialized for serial processing (Paivio & Csapo, 1969; Neisser, 1967). It is necessary to add to this framework a section on reading, that activity in which we have learnt visual symbols as representations of the auditory symbols we know as language. It is because language is first learnt as an auditory/speech system that the visual characteristics of words, letters, and digits are used primarily to refer
back to this system - although it is not known whether it is auditory, articulatory, or abstract (Wickelgren, 1969b). In spite of the prominent part played by language in perception and memory, naming is not necessarily involved in the perception of objects and attributes whereas it cannot be separated from the perception of words.

A detailed analysis of labelling and meaning provides some insights on the amount of information carried by verbal and pictorial stimuli. There is a need for research to determine the conditions under which certain stimulus features are stored and to determine the relative importance and number of features stored. The operation of context is not clearly understood - when a series of items is constant for one attribute it seems unlikely that this attribute will be stored for each item and yet it is probable that a change would be noticed. Perhaps the effect of set is to establish an early check point on attributes expected to be constant - if there is no change on these attributes no information about them will be stored. Such check points could be permanent or temporary. For example, print is usually black on a white background and so the colour of a particular section of print will be stored only if it is not black. This would involve a permanent check point and would be economical in terms of storage as print can always be assumed to be black unless some colour is specifi-
cally stored. During experimental sessions temporary check points are probably set up for the type of material expected.

The study of labelling and meaning also provides an understanding of the great variety of forms of encoding which are possible. One way to investigate these forms experimentally is to vary the stimuli along dimensions such as auditory similarity, physical position of the stimuli, associated sense impressions and so on. Recent work by Wickens, Born, & Allen, (1963), Wickens (1968), Wickens & Clark (1968), and Wickens, Clark, Hull, & Wittlinger (1968), discussed in section 1. 4. 5. has been along these lines. An alternative to the proactive inhibition explanation they offer is one in terms of what Shiffrin & Atkinson (1969) call self-addressable storage. The check points suggested may serve to address storage, different checked values relating to different areas of storage. Therefore a change in the attribute checked would result in storage in a different area and a subsequent search through this other area, so that erroneous recall of previous items would be less likely.

It is a moot point whether all of the richness of meaning is encoded separately or whether there is, for example, one main store and associated programs for the generation of images, associates, etc. This has been dis-
cussed for Experiment 2 where it was suspected that the visual form of words labelling pictorial stimuli was not stored for each item but rather was generated as an image upon recall of the words. However, there do seem to be at least two sorts of storage - verbal and nonverbal, the latter being stored with respect to the sensory modality involved. Factor II in Experiment 2 indicated that verbal and pictorial encoding were alternate forms; the more either was reported the better was recall (Factor I). In this second experiment visual images at recall were not just a cognitive "luxury" as Neisser (1967) would have it. Yet in Experiment 1 visual imagery was nonfunctional. As Sheehan & Neisser (1969) found, the relation between reported imagery and recall depends on the task. They found a functional relation only for incidental learning; comparison of Experiments 1 and 2 indicates that visual imagery is useful only for recall of nonsequential information. It should also be noted that visual imagery occurred less for the task where it was not related to better recall (i.e. Experiment 1). Some people are less addicted to luxuries than others.

These results are across groups or across subjects as are those of Paivio et al. (reviewed in Paivio, 1969). Sheehan & Neisser (1969) found that even when no significant between group or between subjects differences were found there was a relation within individual Ss for higher vividness ratings with designs recalled more accurately.
In summary, it may be said that information can be encoded in many forms. Pictorial and verbal encoding appear to be alternative forms for visual material. The frequency of usage and the effectiveness of the form of encoding depend on the type of task; the more serial the task, the more effective verbal encoding will be. When memory is superior for pictorial material it is probably because both forms of encoding are used. The number of attributes carried by each type of material is also important. Contrary to some current views, visual imagery is not always nonfunctional.

6.3 The use of a questionnaire

Experiments 1 and 2 were the first in which questionnaires were used to investigate differences in encoding for verbal and pictorial stimuli. These questionnaires were more extensive than those used in other memory experiments. In some paired-associate studies (e.g. Yuille & Paivio, 1963) Ss indicated what mediators they used; their descriptions were classified as verbal, imaginal, repetition, other or nothing. Sheehan & Neisser (1969) obtained imagery ratings on the shortened form of Betts' scale but no measure of verbal encoding. Harris & Haber (1963) reported that only 2 of 77 Ss used a visual image for recall of a briefly presented stimulus but it is not clear if Ss were allowed to respond that they used both imagery
and words. Much of the very early work assumed individuals to possess one main form of imagery and to be visiles, audiles, or some such thing. Betts (1909) showed that individuals who were good at one form of imagery tended to be good at other forms as well but there was generally little emphasis on mixed strategies although Davis (1932) did note that recall for Roman and Arabic numberals often involved both visual and auditory imagery.

In Experiments 1 and 2 the results obtained from the questionnaires were very useful for interpreting the results from error measures and there seems no reason why this technique should not be used further. The Zeitgeist is changing again as evidenced by Dulany's (1968) article. However, the relation of imagery in one specific task to general imagery would be better investigated with a standard test such as the shortened form of the Betts' Questionnaire Upon Mental Imagery (Sheehan, 1967) than with one or two questions of the sort used in Experiment 1.

It would also be interesting to repeat, say, Experiment 2 with various sets of instructions. Half the Ss could be informed of the purpose of the experiment and the subsequent questionnaire before beginning any trials. Sheehan & Neisser (1969) found that asking Ss about imagery resulted in higher imagery ratings on subsequent trials; it would be worthwhile to see what effect prior instructions would
have on reported encodings for a task like that in Experiment 2. Also, prior to answering the questionnaire half of the Ss could be given detailed instructions to answer each question separately and to ignore any theories of memory they might have or any ideas about consistency in answering questions. Alternatively, different models of memory could be explained to Ss before they filled in the questionnaire to see if, for example, knowledge of a theory which emphasises verbal encoding affects the levels of verbal encoding reported.

Replications could also be run with instructions to use different techniques of encoding. Mediation instructions have been employed with some success in paired-associate studies (Paivio & Yuille, 1967; Yuille & Paivio, 1968; Schnorr & Atkinson, 1969).

Questionnaires can be useful tools; adequate pre-testing and a greater knowledge of the effects of instructions would make them even more useful in conjunction with tests of memory.

6.4 Other suggested experiments

Apart from the replications suggested above with changes in the questionnaire and encoding instructions, there are two main experiments which would test the conclusions reached about encoding. These are experiments to fill the empty or near empty cells in Table 27:
i) a serial recall experiment with the number of attributes on verbal and pictorial material equated for and

ii) an experiment using more conventional verbal and pictorial material, not equated for the number of attributes, and testing for recall of a single item after interpolated activity (i.e. a replication of Dominowski & Gadlin's third experiment (3.2) but using only words and pictures without numbers).

Variations on Experiment 2 using a less limited stimulus set, a different number of attributes, or recognition instead of recall would clarify a number of details and extend the generality of the results. The template plus blank sheets system used for recognition responses in Experiment 1 took an excessive time to score and methodological or technological advances would be required before replication of this experiment.

Experiments manipulating instead of controlling for the number of attributes encoded from verbal and pictorial stimuli would also be of theoretical interest.

6.5 Conclusions

The diagrammatic model for visual input (2.4) was used to illustrate possible forms of encoding for verbal and pictorial material. Experiments 1 and 2 investigated the relative likelihood of these encodings. The results
of these experiments in conjunction with those from other studies suggest the following conclusions about encoding:

i) the type of encoding depends largely on the task; the more serial the task the more verbal encoding will be used and the less visual encoding will occur (5.6.3),

ii) familiar pictorial material is usually labelled and thus encoded verbally as well as visually,

iii) visual characteristics of words are seldom stored,

iv) images of referents or associates may be aroused by words; these images may be visually encoded,

v) the type of encoding also depends on the rate of presentation as perception, labelling, and image arousal all take time (Paivio & Csapo, 1969).

The amount of verbal encoding appears to be always related to better recall; the amount of visual encoding is related to better recall, at least for nonsequential information.

Pictorial stimuli will be remembered better than verbal stimuli if either or both of the following have occurred:

i) more attributes have been encoded from pictorial than from verbal stimuli,

ii) verbal encoding has occurred to the same extent for both types of stimuli but more visual encoding has occurred for pictorial stimuli.
APPENDIX I

I.1 Types of stimuli

Four types of stimuli were used in Experiments 1 and 2:

i) verbal compact,  
ii) verbal distributed,  
iii) pictorial compact, and iv) pictorial distributed stimuli.

For each type of stimuli there were 27 stimulus slides, the total number of possible combinations of the following attribute levels:

<table>
<thead>
<tr>
<th>Size</th>
<th>Colour</th>
<th>Pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>large</td>
<td>red</td>
<td>dotted</td>
</tr>
<tr>
<td>medium</td>
<td>blue</td>
<td>striped</td>
</tr>
<tr>
<td>small</td>
<td>green</td>
<td>checked</td>
</tr>
</tbody>
</table>

Examples of each type of stimuli are given on the following pages — a large red dotted stimulus is shown throughout.

Four practice slides were constructed for each stimulus type. These slides were:

- large orange wavy
- small yellow plain
- medium pink mottled
- large yellow wavy
a) Verbal distributed

Figure 7

b) Verbal compact
red

large

dotted

large red dotted
I.2 Types of Response Templates

Figure 9a
photograph of a foolscap verbal response template
Figure 3b

photograph of a foolscap pictorial response template
APPENDIX II

II.1 Instructions for Experiment 1

(Numbers in brackets refer to subsequent changes and additions).

"Now, I am going to show you some slides and I want

(1) to see how well you remember them. I will show you two slides
in succession. Your task is then to remember what you have
seen and to mark your answer sheet in the appropriate way.
That is, you see two slides and then you write down what
they were. Then I show another two slides and so we go on.

Write 1 in the hole under the words figure corresponding
(1) to the first slide you saw and 2 under the words figure for the
second slide.

After making your answer, pull your answer sheet out
from under the template - the cardboard. Make sure you pull
out only one sheet and check that there is paper under all
the template holes for the next trial.

The slides will be shown for a short time and then you
will have a limited period of time to answer - about this
(2) long...10-15 secs..."This means that you will have to work
(3) very quickly. Always put down an answer for each slide,
(1) that is two answers each time. If you can't remember a
slide, then guess.

(4)

Any questions?

Now we will have two practice trials using slides which
are just like those in the rest of the session except that they have different colours and patterns. Don't worry if you get muddled -- this is just a practice."

(5) .....presentation of the two practice trials.....

Check Ss' work, answer queries, collect practice response sheets and templates, and issue the templates for the experimental trials.....

"You may have noticed minor imperfections in the slides and in the templates. Don't worry about these. We tried to make everything as precise and perfect as possible, but we just could not do this completely.

so that you can answer quickly please study the template to see how the answers are arranged".....approximately 2 minutes.....

"Right, now we are ready to begin. Remember how to answer and always give an answer for each slide."

..........Present stimuli for all trials.........." "Thank you for enduring this session. Before you go would you please answer this questionnaire."

(1) - From Pilot Study 3 onward, there were three, not two slides per trial.

(2) - From Pilot Study 3 onward, the response time was 20-30 secs., not 10-15 secs.

(3) - "Please look at the screen until all three slides have been presented and on no account try to put down
an answer before you've seen them all. This is a memory task - if you reply before you've seen all the slides you won't really be doing a memory task and I will have to discard your results. Anyway, if you do try to write down the 1st, you'll miss the 2nd slide. Please don't be like some subjects who found the right places with their fingers and then looked to see where their fingers were when all three slides had been presented." These additional instructions were given after three subjects in Pilot Study 3 were found using the finger placement technique.

(4) - "Don't worry if you have to guess often - if everyone got everything right I wouldn't learn much." This instruction was used only in the main experiment.

(5) - For Ss in the pictorial distributed stimuli condition: "The slides you will see show three figures or shapes - the left one tells you size, that is, it is either larger, the same medium size, or smaller than the other two shapes. The middle shape will vary only in colour and the one on the right will vary only in pattern. For example, you might see a slide which was large on the left, red in the middle, and dotted on the right. Each slide will have these three things on it - size, colour, and pattern - and you make one response per slide."
(6) For Ss in the pictorial compact stimulus condition, who were the only ones who actually had to learn what perceptual sizes corresponded to what size labels: "These first three slides will be, 'large, small, and then medium'."

(7) For Ss in the pictorial groups only: "For example, any variations in shape are entirely accidental."

(8) In the main experiment, all Ss were told: "Work independently and don't worry if you take longer than anyone else as long as you are working within the time limit. Don't worry if you drop your pencil or your paper - I'll always wait. I just want to keep things moving at a constant rate for all subjects so that their results are comparable."

II.2 The Questionnaire used in Experiment 1

The figures given in brackets underneath each question are the percentages of the 96 Ss in Experiment 1 who responded with each category.

During Pilot Studies 3 and 4 certain difficulties were found with the questionnaire and so the following additional instructions were given verbally in Experiment 1.

i) For "Units Studied (past and present)" some Ss put down the number of units and not the names of the units. Subsequent Ss were asked for the names of the units they studied.
ii) Question 6 - "What percentage of the time do you think you were right?". This question is ambiguous and so Ss were told, "This question doesn't mean, 'What percentage of the time do you think that you got all three slides right?'; it means, 'What percentage of the slides do you think you got right?' For example, if you usually got only one of the three slides right that would be about 30%; if you mostly got two right, that would be about 60%. Do you understand?"

iii) Question 7 was designed for the two slides per trial task and was nonsensical for the three slide task. The following explanation was given but it seemed unlikely that all Ss really knew how to answer this question and so it was dropped from all analysis:-

"This question is from the time when I showed people only two slides at a time. This was too easy a task so I dropped it. Now, these two slides were always completely different. When I came to make the sets of three slides I made the 2nd completely different from the 1st and the 3rd completely different from the 2nd. Because of the number of alternatives, this meant that the 1st and 3rd slides were sometimes similar and twice they were the same. The question you now answer is whether or not you realised that the 1st and 2nd, and the 2nd and 3rd slides were
always different. Any questions?"

iv) Ss were told to put down any comments they wanted to and to ask any questions if necessary.
QUESTIONNAIRE

Name:________________________

Sex:__________________________Age (in years):____________________

Year at University:____________________

Units Studied (past and present):____________________

Answer 1) before turning the page

1) Write briefly on how you tried to cope with the task in this experiment.
2. Did you, after seeing a stimulus (i.e. a slide) on the screen, think of it in words e.g. "large, red, and dotted"?

Always                  Often              Sometimes          Never
   (41.7%)                (45.8%)           (12.5%)            (0.0%)

3. If you did think of stimuli in words at all, did you "hear" these words in your mind?

Always                  Often              Sometimes          Never
   (39.6%)                (34.4%)           (18.7%)            (7.3%)

4. If you did think of stimuli in words at all, did you think of them:
   a) only in terms of size, colour and pattern e.g. "large, red, dotted"
   OR
   b) with form as well e.g. "a large, red, dotted blob"?

   a)                    b)
   (91.7%)               (8.3%)

5. Did you "see" the stimuli in your mind, either as words or as pictures?

Always                  Often              Sometimes          Never
   (10.4%)                (27.1%)           (40.6%)            (21.9%)

6. What percentage of the time do you think that you were right?

   100% 90% 80% 70% 60% 50% 40% 30% 20% 10% 0%
   (0.0% 1.0% 4.2% 7.3% 26.0% 21.9% 17.7% 11.5% 6.2% 4.2% 0.0%)

7. Did you realize that no pair of stimuli presented on any one trial were ever the same size, colour, or pattern?

Yes                 No
   (65.3%)             (34.7%)
8. Did you keep rehearsing what the stimuli were until you made your answer?

<table>
<thead>
<tr>
<th>Always</th>
<th>Often</th>
<th>Sometimes</th>
<th>Never</th>
</tr>
</thead>
<tbody>
<tr>
<td>(9.4%)</td>
<td>(24.0%)</td>
<td>(42.6%)</td>
<td>(24.0%)</td>
</tr>
</tbody>
</table>

9. Can you remember clearly what things look like?

<table>
<thead>
<tr>
<th>Always</th>
<th>Often</th>
<th>Sometimes</th>
<th>Never</th>
</tr>
</thead>
<tbody>
<tr>
<td>(7.3%)</td>
<td>(45.8%)</td>
<td>(43.8%)</td>
<td>(3.1%)</td>
</tr>
</tbody>
</table>

10. Think of where you had breakfast this morning. Can you "see" the scene and "look over it" with everything being:

   a) as clear as if it were in front of you
   b) clear only in parts
   c) rather dim and hazy
   d) almost non-existent?

<table>
<thead>
<tr>
<th>a)</th>
<th>b)</th>
<th>c)</th>
<th>d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(51.0%)</td>
<td>(35.5%)</td>
<td>(8.3%)</td>
<td>(5.2%)</td>
</tr>
</tbody>
</table>

11. How much are you interested in the visual arts - e.g. painting, sculpture, photography etc.?

<table>
<thead>
<tr>
<th>Extremely</th>
<th>Moderately</th>
<th>A little</th>
<th>Not at all</th>
</tr>
</thead>
<tbody>
<tr>
<td>(19.3%)</td>
<td>(53.1%)</td>
<td>(21.9%)</td>
<td>(5.2%)</td>
</tr>
</tbody>
</table>

12. How much do you sketch, paint, sculpt, or photograph (artistically, not just for snapshots)?

<table>
<thead>
<tr>
<th>A great deal</th>
<th>Moderate amount</th>
<th>A little</th>
<th>Not at all</th>
</tr>
</thead>
<tbody>
<tr>
<td>(3.1%)</td>
<td>(23.9%)</td>
<td>(36.5%)</td>
<td>(36.5%)</td>
</tr>
</tbody>
</table>
APPENDIX III

Relationships between error measures

Experiment 1

A & T

A = number of attribute errors
T = number of slide errors (a slide was counted as wrong if one or more of the three attributes was wrong)

For both A and T, a response was scored correct only if it referred to the right serial position within a trial as well as the right combination of attribute levels.

Let \( n = \) number of slide presentations

\[ = 3 \times \text{number of trials} \]

Then, \( 0 \leq T \leq n, \quad 0 \leq A \leq 3n, \text{ and } T \leq A \leq 3T \)

The following diagram illustrates the bounded zone within which \((A, T)\) could lie.

To test the hypothesis of equidense packing, the following isosceles triangle was drawn and divided into eight segments of equal area. Only seven points lay outside
this triangle and so they were ignored in the analysis. The remaining 89 points were distributed about the central segment such that, with df = 7, $\chi^2 = 93.7$, p<0.001 ($\chi^2_{crit.} = 24.3$). Therefore the null hypothesis of equidense packing may be rejected; there is a strong functional relationship between A and T.

\[ T \& TP \]

TP is the number of slide errors, scored without regard for the reported position in the presentation trio.  
0 < T < n, 0 < TP < n, and TP < T

AP, SP, CP, and PP bear the same relation to A, S, C, and P as TP does to T. The following diagram for T and TP applies to these other pairs too, except that for A and AP the upper bounds are 3n, not n. The shaded area is the zone within which the point (T,TP) may lie.
To test the hypothesis of equidense packing, the area within which the point \((T, TP)\) could lie was divided into triangles of equal area, as shown in the above diagram. Points for the 36 Ss were almost all in the top triangle.

\[
\text{mean } TP = 0.93 \times \text{mean } T
\]

With \(df = 9, \chi^2 = 592.0 \ (\chi^2_{\text{crit.}} = 27.8)\). Thus there is an extremely strong relationship between \(T\) and \(TP\).

Similar analyses were carried out for \(A \& AP, S \& SP, C \& CP,\) and \(P \& PP.\)

**A \& AP**

\[
\text{mean } AP = 0.54 \times \text{mean } A
\]

\(df = 8, \chi^2 = 189.7, p < 0.001\)

There is a strong relationship between \(A\) and \(AP\).

**S \& SP**

\[
\text{mean } SP = 0.54 \times \text{mean } S
\]

\(df = 9, \chi^2 = 63.54, p < 0.001\)

There is a strong relationship between \(S\) and \(SP\).

**C \& CP**

\[
\text{mean } CP = 0.41 \times \text{mean } C
\]

\(df = 9, \chi^2 = 110.3, p < 0.001\)

There is a strong relationship between \(C\) and \(CP\).
mean PP = 0.61 x mean P

df = 9, \chi^2 = 124.9, p < 0.001

There is a strong relationship between P and PP.

In Experiment 1 the range of scores was so wide that the discontinuous nature of each error measure could be ignored.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Range</th>
<th>Measure</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>5,317</td>
<td>AP</td>
<td>5,127</td>
</tr>
<tr>
<td>T</td>
<td>4,150</td>
<td>TP</td>
<td>4,120</td>
</tr>
<tr>
<td>S</td>
<td>0.110</td>
<td>SP</td>
<td>0.108</td>
</tr>
<tr>
<td>C</td>
<td>1,104</td>
<td>CP</td>
<td>0.45</td>
</tr>
<tr>
<td>P</td>
<td>2,103</td>
<td>PP</td>
<td>2.58</td>
</tr>
</tbody>
</table>

This was not so for Experiment 2 where the range of A was (0,17) and the range of T was (0,9).

Experiment 2

As there was only one slide per trial no series position errors could occur. Therefore, only the relationship between A & T required investigation. This was carried out for the 30 seconds group alone as the range for A in the 0 seconds group was a mere (0,3).

A & T

Because both A and T were discrete and varied over small ranges, the restriction T \geq A/3 could not be taken to imply the straight line, T = A/3 as a lower bound. The shaded
isosceles triangle in the following diagram was divided into four equal sized areas and $\chi^2$ calculated ($\chi^2 = 29.8, df = 3, p < 0.001$). Thus the points ($A,T$) are not distributed equally over the total possible area. (10 points lay outside the isosceles triangle, 8 at the apex $(1,1)$, 4 at $(0,0)$, and 1 on the lower boundary line at $(5,2)$). The remaining 73 points were used in the $\chi^2$ calculation.)

The Relationship between $A$ & $T$ in Expt. 2
APPENDIX IV

Treatment of the compact/distributed factor

There is a conceptual problem with the notion of "distributedness" as applied to both verbal and pictorial stimuli. "Distributedness" is more determined by the physical characteristics of the stimuli than, say, "beauty" but less so than "volume" or "colour". This is because the compact forms of verbal and pictorial stimuli may be different in size although the "distributedness" of each type can be ranked by the physical separation of attributes. Compact verbal and pictorial forms are the familiar, conventional representations; distributed forms are obtained by increasing the spacing between attributes for either form. Table 28 shows the average distance between the centres of each attribute over the total width of each slide as a measure of physical spacing. The value for distributed verbal stimuli depends on whether Ss read size, colour, and then pattern or colour, size, and then pattern.

Table 28

<table>
<thead>
<tr>
<th></th>
<th>Verbal</th>
<th>Pictorial</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Compact</td>
<td>Distributed</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Compact</td>
<td>Distributed</td>
</tr>
</tbody>
</table>

* average distance if colour, size, and pattern are read in that order instead of size, colour, and pattern.
The following factors occurred in Experiments 1 and 2.

\[ E = \begin{cases} 
A - \text{verbal/pictorial stimulus factor} \\
B - \text{compact/distributed factor} \\
C - \text{verbal/pictorial response factor in Experiment 1} \\
\quad \text{silence/voicing factor in Experiment 2} \\
D - \text{male/female factor} 
\end{cases} \]

<table>
<thead>
<tr>
<th>( a_1 ) - verbal stimuli</th>
<th>( a_2 ) - pictorial stimuli</th>
</tr>
</thead>
<tbody>
<tr>
<td>( b_1 ) - compact</td>
<td>( b_1 ) - compact</td>
</tr>
<tr>
<td>( b_2 ) - distributed</td>
<td>( b_2 ) - distributed</td>
</tr>
</tbody>
</table>

Given that it is uncertain as to whether compact/distributedness is the same factor for both verbal and pictorial stimuli, there are three possible types of analysis of variance:

i) combining factors A and B into one factor E with four levels,

ii) treating all factors as independent,

iii) treating factor B as nested under factor A.

Table 29 shows what interactions are lost with each type of analysis. Note that in i) and iii) the conceptual structure of \( b_1, b_1 \) and \( b_2, b_2 \) is lost for there are four separate groups, \( e_1, e_2, e_3, e_4 \), or four nested groups, \( b_1, b_2, b_3, b_4 \), respectively.
Table 29
Sums of Squares for ANOVA
(from Winer, 1962)

<table>
<thead>
<tr>
<th>(i) 3 factors</th>
<th>(ii) 4 factors</th>
<th>(iii) B nested under A</th>
</tr>
</thead>
<tbody>
<tr>
<td>E = A+B+AB</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>-</td>
<td>B</td>
<td>Bw.A = B+AB</td>
</tr>
<tr>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>-</td>
<td>AB</td>
<td></td>
</tr>
<tr>
<td>EC = AC+BC+ABC</td>
<td>AC</td>
<td>AC</td>
</tr>
<tr>
<td>BD = AD+BD+ABD</td>
<td>AD</td>
<td>AD</td>
</tr>
<tr>
<td>-</td>
<td>BC</td>
<td>(Bw.A)C = BC+ABC</td>
</tr>
<tr>
<td>-</td>
<td>BD</td>
<td>(Bw.A)D = BD+ABD</td>
</tr>
<tr>
<td>CD</td>
<td>CD</td>
<td>CD</td>
</tr>
<tr>
<td>-</td>
<td>ABC</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>ABD</td>
<td></td>
</tr>
<tr>
<td>ECD = ACD+BCD+ABCD</td>
<td>ACD</td>
<td>ACD</td>
</tr>
<tr>
<td>-</td>
<td>BCD</td>
<td>(Bw.A)CD = BCD+ABCD</td>
</tr>
<tr>
<td>-</td>
<td>ABCD</td>
<td></td>
</tr>
</tbody>
</table>

Method i) results in the most loss of information.
Method ii) makes no unjustifiable assumptions but ignores the b1, b1 and b2, b2 structure. Method ii) assumes that factor B may be treated as a separate factor but if any of the interactions involving both factors A and B are significant it is unclear if this is because B ≠ B' or if there is a genuine interaction. Furthermore, even if an interaction is
nonsignificant this could be because there was none or because the inequality of $\beta$ and $\beta'$ and the interaction were antithetical. However the prior probability of this seems so low that the simple explanation of no interaction may be accepted.

Thus, given that under method ii) there are no interactions involving both factors A and B it would seem reasonable to proceed with that method of analysis. If, however, any of these interactions are significant the interpretation is unclear and methods ii) and iii) are equally useful.
APPENDIX V

V.1 Instructions for Experiment 2

Necessarily there were slight differences in the instructions for the different groups. These are sometimes indicated by superimposed alternatives; otherwise -

{ } - indicates extra instructions for pictorial stimulus groups,

[ ] - extra instructions for distributed stimulus groups,

( ) - extra instructions for groups who voice at presentation,

- extra instructions for 30 seconds groups.

"Are you colourblind?"

In this Experiment there will be two tasks - a memory and a counting task.

The counting task is one in which you will see a three figure number on the screen. I want you to say the number out loud and then to count backwards from it in threes e.g. if you saw the number three hundred and seventy you would say, 'three hundred and seventy, three hundred and sixty seven, three hundred and sixty four, three hundred and sixty one, three hundred and fifty eight and so on.' Please say the numbers in full - 'three hundred and sixty eight.' You must try to count accurately. We will practice this. I will show you a three figure number on the screen, you say it out loud, and then count backwards in threes out loud. The metronome will tick every second and you should aim at
counting at that rate. "Ready"...30 seconds...."Stop."

"You've probably noticed the pattern by now - if you start with, say, a number ending in zero you go nought, seven, four, one, eight, five, two, nine, six, three, and nought again on the last digit".

Correct any errors made in practice.

"We will have one more practice. "Ready"...30 seconds...

Correct any more mistakes.

0 seconds groups: "The second task is quite a simple one - you will see a slide with words on it, telling you something about size, colour, and pattern. (When you see this slide say out loud what is on it). When I second later you hear this tone........write down in words what was on the slide, that is size, colour, and pattern. Size, colour, and pattern are equally important. You will have 20 seconds to answer and then you will see a three figure number and count backwards until you hear this tone again. Then I will say 'ready' and another trial will begin. When I say 'ready' it just means 'look at the screen'."

30 seconds groups: The memory task is quite a simple one - you will see a slide with words on it, telling you something about size, colour, and pattern. (When you see this slide say out loud what is on it.) Then you will see a slide with a number on it and you'll count backwards as you practiced until you hear this tone.......which means write down in words what was on the first slide, that is
size, colour, and pattern. Size, colour, and pattern are equally important. You will have 20 seconds to answer and then another trial will begin. When I say 'ready' it just means 'look at the screen'."

"When you are counting, concentrate only on that; when writing down a slide don't practice counting."

So that you will know what the slides you have to write down look like I will show you some but these will have different colours and patterns from the ones you will see in the experiment itself. {All slides are supposed to show the same shape[s] so ignore all minor variations.}

Only size, colour, and pattern matter. {[The slides will show 3 blobs - the left blob tells you size - if it's larger, the same, or smaller than the other two blobs which will always be the same size - the middle blob tells you colour, and the right blob tells you pattern.]} I'll show you 3 slides, each one for longer than in the experiment so that you can have a good look at it. (Say out loud what you see each time.) These are just to look at, not to write down remember .......

{This one is large....this one is small....this one is medium.}

The 4th one you will see just for 1 second which is the length of time you'll have in the experiment.........

Now, the slides you will see won't have those colours or patterns. Have a good look at this chart which shows that only three sizes, three colours, and three patterns are
possible.

We are ready to begin. Please cover up your answers as you go with this chart. The order will be:

30 seconds groups: "slide to remember (and you must say it out loud), the number slide and counting until the beep, and then recall of the first slide."

0 seconds groups: "-slide to write down (and you must say it out loud), writing down what was on the slide, then the number and backwards counting."

"If you think of a way to cheat, please don't."

"Ready"..........................
V.2 Questionnaires

Questionnaire for the 30 seconds groups

QUESTIONNAIRE

Name: ____________________________________________________________________ Age (in years): __________

Sex: ____________________________________________________________________ Faculty: __________________________________________________________________

Year at Varsity: __________________________________________________________________

Fulltime or parttime?


In the following questions the words
Never  Sometimes  Often  Always
will be used. Please use these words as if they corresponded
to 0-10%  10-50%  50-90%  90-100%
Perhaps the words "almost never" and "almost always" should
be used but they are rather clumsy to use often.

Add explanatory comments to questions if you feel they
will illuminate your answers.
Task 1 Memory

When you saw the slide you had to remember

1. Did you think of it in words e.g. "large red dotted"?
   Always          Often          Sometimes        Never
   (52.1%)         (17.7%)       (16.7%)         (13.5%)

2. If you did think of the slide in words at all, did you "hear" these words in your mind?
   Always          Often          Sometimes        Never
   (27.1%)         (15.6%)       (25.0%)         (32.3%)

3. If you thought of the slide in words at all did you think of it
   (82.3%) a) only in terms of size, colour, and pattern e.g. "large red dotted"
   (17.7%) b) or with form as well e.g. "large red dotted blob"?

4. Did you manage to repeat the slide to yourself before you had to start counting backwards?
   Yes              No
   (55.2%)          (44.8%)

5. If you did do so, how many times on the average would you have repeated it, not including the first time you perceived the slide.

6. While counting backwards did you repeat the slide to be remembered to yourself?
   Always          Often          Sometimes        Never
   (1.0%)          (3.1%)        (33.4%)         (62.5%)

7. While counting backwards did you visualize the slide you had to remember?
   Always          Often          Sometimes        Never
   (2.1%)          (2.1%)        (30.2%)         (65.6%)
When trying to remember the slide you had seen

8. Did you remember it in words?
   Always   Often   Sometimes   Never
   (35.4%) (25.0%) (20.3%) (18.8%)

9. Did you seem to "hear" what it was?
   Always   Often   Sometimes   Never
   (6.3%) (22.9%) (30.2%) (40.6%)

10. Did you see an image of words?
    Always   Often   Sometimes   Never
    (3.1%) (10.4%) (27.1%) (59.4%)

11. Did you see an image of a picture?
    Always   Often   Sometimes   Never
    (15.6%) (19.8%) (30.2%) (34.4%)

12. Did you treat size, colour, and pattern
    a) as equally important
    OR
    b) concentrate on ______________________ (State which)

13. What percentage of the time do you think that you remembered correctly?

    Size  0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%
    Colour 0 10 20 30 40 50 60 70 80 90 100
    Pattern 0 10 20 30 40 50 60 70 80 90 100

14. What percentage of the time do you think that you remembered everything on the slide correctly? (don't try to do fancy calculations from your previous questions, just state the figure which is closest to how you feel you did)
    0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

15. Do you think that you muddled the slide you had just seen with the previous one?
    Always   Often   Sometimes   Never
    (0.0%) (13.5%) (54.2%) (32.3%)
16. If there is anything else about your attempt to remember which seems important please state it briefly here.

17. Do you think that your rate of counting improved over the trials?

Yes
(26.0%)

No
(74.0%)

18. Backwards counting seems to be an activity which does not lend itself to introspection but if you have any further comments please state them here.
Questionnaire for the 0 seconds groups

QUESTIONNAIRE

Name: ____________________________ Age (in years): ______________________

Sex: ____________________________ Faculty: ____________________________

Year at Varsity: ______________________

Fulltime or Parttime?

In the following questions the words

Never        Sometimes        Often        Always
will be used. Please use these words as if they corresponded
to 0-10%       10-50%        50-90%        90-100%

Perhaps the words "almost never" and "almost always" should
be used but they are rather clumsy to use often.

Add explanatory comments to questions if you feel they
will illuminate your answers.
Task 1 Memory

When you saw the slide you had to remember

1. Did you always think of it in words e.g. "large red dotted"?
   
   Always   Often   Sometimes   Never
   (65.6%) (16.7%) (8.3%)  (9.4%)

2. If you did think of the slides in words at all, did you "hear" these words in your mind?
   
   Always   Often   Sometimes   Never
   (40.6%) (24.0%) (13.5%) (21.9%)

3. If you thought of the slide in words at all did you think of it
   
   (89.6%) a) only in terms of size, colour, and pattern
   e.g. "large red dotted"
   (10.4%) b) or with form as well e.g. "large red dotted blob"?

4. Did you manage to repeat the slide to yourself before the tone sounded?
   
   Yes   No
   (70.8%) (29.2%)

5. If you did do so, how many times on the average would you have repeated it, not including the first time you perceived the slide?

6. Did you find that although you had seen the slide clearly you could not remember what was on it when you came to write your answer?
   
   Always   Often   Sometimes   Never
   (0.0%) (1.0%) (12.5%) (96.5%)

7. Do you think that you muddled the slide you had just seen with the previous one?
   
   Always   Often   Sometimes   Never
   (0.0%) (0.0%) (3.1%) (96.9%)
8. If you have any other comments state them here.

Task 2 - Backwards Counting

9. Did you find that counting became easier over the trials?

Yes          No
(21.98)      (78.18)
APPENDIX VI

Labelling used in Experiment 2

Ss who saw pictorial stimuli were not told what words to use to describe these stimuli. Hence, an analysis was carried out to see what labels were used. The following tables show the number of people using each word; where more than one word was used by the same S the more frequent one was counted. Results from the 0 and 30 seconds groups were combined as there were no differences between them.

Table 30

Size labels

<table>
<thead>
<tr>
<th></th>
<th>Large(x)</th>
<th>Big(ger)</th>
<th>Medium</th>
<th>Same</th>
<th>Average</th>
<th>Small(er)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vc vd</td>
<td>96</td>
<td></td>
<td>96</td>
<td></td>
<td></td>
<td>96</td>
</tr>
<tr>
<td>Pc</td>
<td>48</td>
<td>44</td>
<td>43</td>
<td>28</td>
<td>19</td>
<td>2</td>
</tr>
<tr>
<td>Pd</td>
<td>44</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>48</td>
</tr>
</tbody>
</table>

V - verbal stimuli  
P - pictorial stimuli  
C - compact stimuli  
D - distributed stimuli

There was little difference in the words used for verbal and pictorial stimuli except for the middle size with pictorial distributed stimuli. Actually size words were used in the instructions to the pictorial stimulus groups.

For colour, everyone used the three words, "red", "blue", and "green" although one or two Ss in the pictorial stimulus
groups occasionally qualified "green".

**Table 31**
Pattern Labels

<table>
<thead>
<tr>
<th></th>
<th>Dotted</th>
<th>Spotted</th>
<th>Speckled/Stippled</th>
<th>Striped/Stripes</th>
<th>Lines</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_C &amp; V_D$</td>
<td>96</td>
<td></td>
<td></td>
<td>96</td>
<td></td>
</tr>
<tr>
<td>$P_C$</td>
<td>27</td>
<td>21</td>
<td></td>
<td>15</td>
<td>18</td>
</tr>
<tr>
<td>$P_D$</td>
<td>19</td>
<td>23</td>
<td>6</td>
<td>15</td>
<td>13</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Diagonal Stripes or lines</th>
<th>Checked</th>
<th>Squared/squares</th>
<th>Crisscrossed/crossed</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_C &amp; V_D$</td>
<td></td>
<td>96</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P_C$</td>
<td>15</td>
<td>13</td>
<td>21</td>
<td>14</td>
</tr>
<tr>
<td>$P_D$</td>
<td>20</td>
<td>8</td>
<td>28</td>
<td>12</td>
</tr>
</tbody>
</table>

Thus many Ss did not use the words "striped" or "checked" to describe the pictorial equivalents of these verbal stimuli.
REFERENCES


ARCHER, J.E. 'Concept identification as a function of obviousness of relevant & irrelevant information.' J. exp. Psychol., 1962, 63, 615-620.


BARRICK, H.P., & BOUCHER, E. 'Retention of visual and verbal codes of the same stimuli.' J. exp. Psychol., 1968, 73, 417-422.


CONRAD, R. 'An association between memory errors and errors due to acoustic masking of speech.' Nature, 1962, 193, 1314-1315.


DALLETT, K. & WILCOX, S.G. 'Remembering pictures versus remembering descriptions.' Psychon. Sci., 1968, 11, 139-140.


EPSTEIN, W., ROCK, I., & ZUCKERMAN, C.B. 'Meaning and familiarity in associative learning.' Psychol. Monogr., 1960, 74, No.4 (Whole No.491).


FRAISSE, P. 'Le temps de reaction verbale.' Année Psychologique, 1964, 64, 21-46.

FRAISSE, PAUL. 'Motor and verbal reaction times to words and drawings.' Psychon. Sci., 1966, 12, 235-236.


GREGSON, R.A.M. Personal communication, 1968.


JUNG, C.G. 'Studies in word-association.' London: Heinemann, 1918.


LOESS, H. 'Short-term memory, word class and sequence of items.' J. exp. Psychol., 1967, 74, 556-561.


MORRIS, C. 'Signs, language, and behaviour.' New York: Prentice-Hall, 1946.


MURDOCK, B.B. JR. 'Decoding as a function of the number of bits per chunk.' J. exp. Psychol., 1968, 78, 1-7.


MURRAY, D.J. 'Vocalization-at-presentation and immediate recall, with varying presentation rates.' Quart. J. exp. Psychol., 1965, 17, 47-56, (a).


MURRAY, D.J. 'Vocalization at presentation and immediate recall with varying recall methods.' Quart. J. exp. Psychol., 1966, 18, 9-18, (a).


OSGOOD, C.E. 'Meaning cannot be \( r_m \).' J. verb. Learn. 


PAIVIO, A., YUILLE, J.C., & ROGERS, T.B. 'Noun imagery and meaningfulness in free and serial recall.' J. exp. Psychol., 1969, 72, 509-514.


PETERTSON, L.R., & GENTILE, ANTOINETTE. 'Proactive interference as a function of time between tests.' J. exp. Psychol., 1965, 70, 473-479.

PETERTSON, L.R., & PETERTSON, M.J. 'Short-term retention of individual verbal items.' J. exp. Psychol., 1959, 58, 183-193.


POSSNER, M.I., & KEELE, STEVEN. 'Decay of visual information from a single letter.' Science, 1967, 158, 137-139.

POSSNER, M.I. & KONICK A.P. 'Short term retention of visual and kinaesthetic information.' Organizational Behaviour and Human Performance, 1966, 1, 71-86.

POSSNER, M.I., & ROSSMAN, ELLEN. 'Effect of size and location on informational transforms upon short-term retention.' J. exp. Psychol., 1965, 70, 496-505.


PRIEST, K.F., & CROWTHER M. Personal communication, 1969.

QUINE, W.V. 'The problem of meaning in linguistics.' In W.V. Quine; From a logical point of view, pp 47-64. Cambridge, Mass.: Harvard Univ. Press, 1953.


REICHEL, K.F. 'Age and cultural differences as determinants of word associations: Suggestions for their analysis.' Psychol. Rep., 1965, 16, 75-78.


RUSSELL, P. N. 'Cross Tabulation - An IBM 360/44 FORTRAN IV program for the construction of contingency tables and calculation of Chi-Square and contingency coefficients.' Behav. Sci., 1969, 14, 166-167.

SAMPSON, J.R. 'Further study of encoding and arousal factors in free recall of verbal and visual information.' Psychon. Sci., 1969, 16, 221-222.


SPERLING, G. 'The information available in brief visual presentations.' Psychol. Monogr., 1960, 74, No.11 (Whole no. 498).


VAUGHAN, G.M. Personal communication to P.M. Russell, 1969.


VIGOTSKY, L.S. 'Thought and speech.' Psychiatry, 1939, 2, 29-54.
WALLACE, W.P. 'False recognition produced by laboratory established associative responses.' Psychon. Sci., 1967, 7, 139-140.

WAUGH, N.C., & NORMAN, D.A. 'Primary memory.' Psychol. Rev., 1965, 72, 89-104.


WICKELGREN, W. A. 'Short-term memory for phonemically similar lists.' Amer. J. Psychol., 1965, 78, 567-574, (d).


WILLIAMS, H.L., BEAVER, W.S., SPENCE, M.T., & RUNDELL, O.H. 'Digital and kinaesthetic memory with interpolated information processing.' J. exp. Psychol., 1968, 80, 530-536.


WOOD, GORDON. 'Higher order memory units and free recall learning.' J. exp. Psychol., 1969, 80, 286-288.

