THE PROCESS OF REMEMBERING PICTURES

Jean Elizabeth Bird

A thesis submitted in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy
of the Australian National University

May, 1976

This thesis describes original research carried out by the author during the tenure of a Commonwealth Postgraduate Research Award in the Department of Psychology of the Australian National University from February, 1973, to May, 1976.

J E Bird

J.E. BIRD

ACKNOWLEDGEMENTS

I am very grateful to the many people who have contributed their time and efforts to this research. My supervisor for the project, Dr. Michael L. Cook, has been invaluable in providing new ideas and thoughtful criticisms of the work as it has progressed. His constant interest, and the painstaking care with which he helped me to produce a finished version of the manuscript, are very much appreciated.

I also wish to thank the other postgraduate students and members of the technical and academic staff in the Department of Psychology who have given valuable suggestions and technical assistance in designing experimental apparatus, typing response sheets and proofreading the manuscript. The constructive suggestions made by Ms Norma Chin for the presentation and typing of the thesis deserve special commendation.

Finally, I am indebted to the undergraduate students in the Department of Psychology at the Australian National University who volunteered for my experiments, since without their participation this research could not have been possible.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Eight experiments were carried out to investigate processes involved in remembering pictorial stimuli. Because of difficulties encountered in measuring memory for this kind of material, previous investigations in the area (reviewed in Chapters One and Two) have not successfully estimated the capacity of memory for pictures or specified the nature of its encoding processes. A new recognition paradigm, the exclusion set method, was developed in order to measure the accuracy of the subject's memory for representational drawings. It was used to refute two hypotheses: (1) that memory for pictures has a phenomenally high capacity (Experiment 1), and (2) that encoding of pictures is an automatic process (Experiments 2 and 3). Another method of recognition testing was then devised to examine the constructive nature of the encoding process in more detail, in a study of memory for abstract shapes (Experiments 4 to 8).

Experiment 1 tested subjects' memory for pictures after delays of one, two, seven and 60 days. There was evidence that memory is limited in capacity, since subjects were not accurate on the recognition task. Memory declined after one week's delay and again after two months.

To determine whether the elaboration of subjects' encoding strategies affects memory for pictures, intentional and incidental instructions were manipulated in Experiments 2 and 3. Three instructional groups were tested in each experiment: (1) an Incidental group not told of the memory task and instead given a picture-classification orienting task, (2) a Control group given the orienting task but told about the memory requirement, and (3) an Intentional group allowed to view the pictures freely without an orienting task. In Experiment 2, no differences between any of the groups was found, suggesting that despite subjects' different encoding strategies, all had encoded the same amount of information from the pictures. In Experiment 3, the Intentional group was altered to include a practice trial for this group only.

Intentional subjects given practice recognized more pictures than Incidental or Control subjects. This showed that different encoding activities carried out during presentation of a list of pictures can be variable and under the subject's control, rather than an automatic registration of information into memory.

Since encoding did appear to be a function of the subject's encoding activities, it was hypothesized that recognition would be affected by the amount of time subjects were given to process each picture. In particular, it was predicted that temporal variables would affect subjects' tendency to encode only parts of a stimulus without adequately encoding their combinations. The presence of inter-picture confusions in previous studies suggested that a "fragmented memory effect" is a common outcome of subjects' inadequate encoding activities. Experiment 4 found no effect of longer presentation time or ISI on recognition of shapes, but the fragmented memory effect was demonstrated empirically. The methodology of this experiment was improved and it was found that five seconds of presentation time led to more accurate recognition than two seconds (Experiment 5), though increasing the ISI from 1.5 to seven seconds had no effect (Experiment 6), and that subjects given two seconds of either presentation time or ISI performed more accurately than subjects 0.5 seconds (Experiment 7). Presentation time was found to be more beneficial than ISI, given the same total time. Thus, memory for whole shapes improved with longer presentation times and also with longer ISI's of short duration. A final experiment (Experiment 8) was carried out to see whether the fragmented memory effect could be altered, during fast presentation of pictures, with different perceptual strategies. Strategies altered the attention to different parts of the shapes but did not affect fragmented memory. The encoding process was discussed as a sequence of acts of attention to parts of pictures followed by an integration which is not always successful.

CHAPTER ONE

PROBLEMS IN STUDYING MEMORY FOR PICTURES

1.1 THE NARROW SCOPE OF PRESENT CONCEPTUALIZATIONS OF MEMORY

Theoretical conceptualizations of memory proliferate, and yet the empirical data on which these are based, specifically experiments on verbal learning, pertain to a narrow range of the situations actually encountered in experience. For example, a large collection of modern information-processing models of memory [Norman, 1970] does not contain any theory which relies substantially on evidence other than that provided by our knowledge of language or findings of studies in which subjects learn lists of verbal items. This would be permissible if the experiments designed to study memory were a representative sample of memory situations, or if conclusions could be generalized so that they have relevance for the entire phenomenon of memory. At present, however, there is insufficient evidence to demonstrate the general applicability of conclusions based on studies of verbal learning to the field of memory in general. Until quite recently, materials used in memory experiments have been limited to lists of words, letters, digits and nonsense syllables. Theories of memory have been and are still being generated on the basis of data from studies investigating specifically verbal memories and employing a narrow range of paradigms. There is still a tendency to overgeneralize the trends found in these studies.

Memory for pictures does not fit easily into most current conceptualizations of memory. Typically, these models postulate a sensory register, short-term store and long-term store in memory [e.g., Norman and Rumelhart, 1970; Waugh and Norman, 1965]. The short-term store is usually conceived as a limited-capacity "buffer" in which items are retained by verbal rehearsal. Thus, stimuli must be named in order to be remembered in the system. Preliminary evidence from studies of memory for pictures shows that hundreds of pictures can be recognized after brief, massed presentation, whether testing is immediate or delayed by several days [e.g., Nickerson, 1965, 1968; Standing, 1973]. These data are difficult to reconcile with a limited-capacity short-term store of five to seven items which is separate from a longer term store. There is also disagreement in the literature as to whether picture rehearsal is nonverbal [e.g., Haith, 1971] or whether pictures can be rehearsed at all [e.g., Shaffer and Shiffrin, 1972]. In fact, there is direct experimental evidence to suggest that pictures are not remembered in the same way as words. For example, we have a higher memory capacity for pictures than words [e.g., Lieberman and Culpepper, 1965; Paivio, Rogers and Smythe, 1968] or verbal descriptions of scenes [Dallett and Wilcox, 1968]. Also, serial position curves for pictures are different even when recall is carried out verbally for both pictures and words [G. Cohen, 1972]. While there are at present innumerable modifications which might overcome any particular criticism [e.g., G. Mandler, 1975], the general criticism remains that these models are based essentially on data from verbal learning experiments, and consequently they are limited in scope.

Recently more sophisticated "semantic" models [e.g., Kintsch, 1972; Rumelhart, Lindsay and Norman, 1972] have defined the substance of memory in terms of "concepts" and their relationships, rather than as words. But these conceptualizations are still based on linguistic analysis of a formal, sequential nature and are not designed to handle the spontaneity and complexity of everyday experience. The purpose of this kind of linguistic analysis has been to expand the idea that speech is represented internally by a deep structure, in which there is no one-to-one relationship between ideas and words, as well as by a surface structure. With their emphasis on language, these models, like verbal models, provide only a limited account of memory.

1.2 MODELS OF MEMORY BASED ON DATA FROM NONVERBAL STUDIES: THEIR PROBLEMS

A few memory models have widened the basis from which they have drawn their experimental support and have attempted to incorporate data from studies of memory for pictures, along with traditional verbal data, into the framework of an information-processing approach. Recently several theorists have discussed memory in terms of two modal stores, a verbal store and a nonverbal, pictorial store. There is a great deal of experimental evidence for the notion that there is short-term memory storage which is not verbal in nature and which is not the same as the very short-term visual store first demonstrated by Sperling [1960] and called "iconic memory" by Neisser [1967]. This visual storage is present for intervals longer than the one to five seconds suggested for iconic memory and is unaffected by visual masking. In a series of experiments discussed by Posner [1969], evidence was presented favouring the existence of a purely visual storage mechanism based on recognition-reaction-time tasks. In these tasks a single letter is presented to subjects, followed by a second letter after a short interval. It has been shown that a stored letter can be matched more rapidly with a physically identical letter (e.g., A followed by A) than it can with a letter having only the same name (e.g., A followed by a). If memory for letters consists of only a verbal representation of the letter name, it should not make any difference whether the second letter is physically the same as the first or not. Since matches based on physical shape are faster than matches based on the names of the letters, Posner hypothesized that matches could be made on the basis of information in visual storage prior to the naming process. Following from this, Coltheart's [1972] model of visual information processing suggested that stimuli are first incorporated into high-capacity iconic memory, then into a fast visual code of low capacity and finally into a slower name code of moderate capacity. Long-term memory could be based either on a verbal or a visual code. Cermak [1971] found that complex, multidimensional free-form shapes exposed for five seconds could be recognized significantly better than chance for up to 20 seconds after they had been presented. The shapes were difficult to label, and it is doubtful whether subjects could have used verbal memory for remembering them.

Some information-processing models have attempted to incorporate recent findings about nonverbal memory into more traditional approaches to memory. Several theorists [e.g., Atkinson and Shiffrin, 1968; Paivio, 1971] have argued that there are two or more short-term stores, and specifically include a short-term verbal store as well as a a short-term visual or pictorial store. Atkinson and Shiffrin [1968] suggested that memory stores have modal "compartments", so that there might be different sensory, short-term and long-term stores for each modality such as the auditory-verbal-linguistic mode and the visual mode. Paivio's [1971] theoretical notions went much further on the basis of later experimentation with pictures. His own work pioneered a great deal of research into memory for pictures using experimental paradigms adapted from verbal learning. His postulation of two separate stores in memory was based largely on the finding that subjects have a higher capacity for pictures than words in memory. Paivio hypothesized that a dimension of abstractness-concreteness was the most important stimulus variable in memory, and that pictures are remembered better than words because of their greater concreteness. He also postulated that stimuli which could be "dually encoded" with both visual and verbal information would be best remembered. This offered an explanation for differences between abstract and concrete nouns: abstract nouns, defined as being more difficult for evocation of images, are remembered more poorly than concrete nouns, which subjects rate as high in image-evoking ability. Paivio's account is the only one to deal with memory for pictures, and it sets his ideas of memory apart from those of verbal learning theorists. The danger with Paivio's approach is that modality-specific information can only have importance in the early stages of perception or the latest response stages of processing. The most important functions in memory are carried out in the association cortex which is not modality-specific in the physical sense. It is likely that memory is abstracted, semantic knowledge, which has access to modality-specific information, as suggested by the semantic network theorists.

Although the wider range of stimulus materials used by Paivio is encouraging, his approach seems destined to lead to more models with a proliferation of memory stores, each based on studies using modality-specific experimental materials. It might be more fruitful to discuss general memory processes which occur in particular situations, with particular materials, depending upon the varied skills subjects bring to the experimental setting.

Another theoretical approach to nonverbal memory is found in the pattern recognition literature. Although primarily directed to perception rather than to memory, this view is relevant here, insofar as pattern recognition is seen as involving a comparison of patterns perceived with stored representations of previously perceived patterns. Sutherland's [1968] discussion is a typical and influential example [see also its development by Reed, 1973]. It postulates that in perceiving, abstract structural descriptions are derived and that recognition consists in matching these with stored structural descriptions of previously encountered objects. Sutherland's account is of interest here in providing (albeit very schematically) a discussion of the nature of visual memory representations. However, no real attention is paid to the questions of encoding strategies, retrieval mechanisms, and capacity for stored information, which must be considered in any theoretical account of memory.

One of the few traditional memory approaches which is process-related and not stimulus-dominated is the "depth of processing" framework [Craik and Lockhart, 1972]. Theorists of this view study words largely as a convenience, and hypothesize that memory is increased by sophistication and elaboration of thought processes concerning the stimulus during its encoding. Encoding does not rely on naming or the semantic analysis of language, though these processes have been used to control the amount of encoding elaboration. The approach is flexible enough to include studies of memory for different kinds of stimuli and different paradigms. Essentially this is because the model is not concerned with the structure of memory storage. Rather, the functions of information stored in a particular way for retention and retrieval are of greatest concern. This allows the model to be testable in a larger range of memory situations than has been possible with previous models, which have had greater formality though they have had less flexibility and generality.

1.3 THE CHOICE OF STIMULI

The most appropriate stimuli to study in memory experiments are real objects, but it is difficult to design well-controlled experiments to study such material. This would be extremely difficult if presentation times for stimuli and the amount of movement and tactile exploration the subject is allowed were to approximate those of "normal" perception. Another problem which would be encountered in studying memory for real objects is the lack of a reliable and valid method for measuring memory. The verbal learning experiments employ highly simplified methodologies which are adequate for studying memory for words. However, these are not adequate for studying other kinds of memory. In particular, it is not possible to measure recall simply by counting the number of items named correctly. The visual, tactile, verbal and other knowledge that subjects have about a stimulus is not measured accurately by asking the subject simply to report on his memory. Clearly, the subject has recall for nonverbal material, but this is a private, mental event (e.g., an image) which may not be communicated to another person directly. This problem, largely ignored in verbal learning studies, has received more attention in studies of memory for pictures.

Pictures as Stimuli

The reasons for a bias towards verbal learning experiments in the study of memory are easily understood. Words are well-learned units, symbols which can easily be understood by others if they are spoken or written as marks on a paper. The exact configuration of lines in letters or numbers is not important. A letter, number or word is taken by the subject as a token of a class of patterns, and style of handwriting, typefont and case are not usually important in remembering this class. Memory experiments have become quite sophisticated in controlling word stimuli. For example, the frequency of usage of a word is a good indication of its inherent familiarity, and the number of letters or syllables it contains may be controlled to make lists of similar items. Similarly, the number of words correctly recalled provides a reliable measure of memory. Pictures, on the other hand, are much more difficult to control. Like words, pictures are marks on a surface which are symbolic. However, the relationship between a picture and its referent in the world is not generally based on an arbitrary convention, as it is in language. Photographs, for example, have correspondence with items in the physical world (or at least their retinal images). Sentences are interpreted syntactically, using formal rules of language, while pictorial conventions do not provide a rigid system for interpretation [Kolers, 1973].

In the present context the term "picture" is used to denote a photograph of an object or a drawing. These differing stimuli can be called pictures because of two essential characteristics:

(1) they are nonverbal; i.e., they are not direct representations of words, letters or digits, and,

(2) each picture is a "physical surface which has been treated or processed or acted upon in such a way that the light causes a perception of something other than the surface itself" [Gibson, 1960].

There are some visual patterns which are neither verbal nor pictorial in the same sense as is a representation of an object. Patterns of dots or lines, ink blots and randomly-generated polygons have also been studied as a nonverbal analogue of the nonsense syllable. These patterns are often studied in place of meaningful pictures (i.e., those with concrete references) because they are easier to control and can be generated in large quantities. In the sense of Gibson's [1960] definition, these patterns are not pictures unless the perceiver attributes some meaning to them other than that they are marks on a surface. There is little doubt that subjects do attempt to attribute meaning to patterns even if it is an idiosyncratic meaning. The use of Rorschach patterns attests to this. Thus there would not seem to be a necessary distinction between pictures with referents in the physical world and those more difficult to interpret. There may be some differences in memory for pictures which differ in meaningfulness, but results of experiments using realistic scenes could be verified in better-controlled experiments using "nonsense" material in a complete investigation of a particular problem.

There is an important reason why pictures are more typical of objects encountered in the real world than are words. Each picture is a unique configuration of lines (with a few exceptions) never encountered in experience before. Words, on the other hand, have been encountered before, and are part of our knowledge of the language. Indeed, many theorists [e.g., Kintsch, 1970] suggest that we remember words simply by "tagging" their pre-existing entries in the linguistic memory store. It is implausible to hypothesize anything similar for pictorial memory.

No doubt it is misleading to think of figurative pictures simply as controlled representations of the real world which can be used to study "real" memory experimentally. Pictures differ from real objects in an important way. Pictures may represent real-world scenes by their resemblance to them, but at the same time they also have a more basic level of perceptual interpretation in that they are the category of objects which can be called "marks on a surface". A painting is a piece of canvas with paint stuck on it, on the first level of interpretation, and a representation of something else on the second level of interpretation.

Pictures can be controlled more easily than other physical objects in order to be used in experiments, but they must not be used exclusively in place of the "real thing". At present there are too many difficulties to overcome in order to study memory for events in the real world. The study of pictures broadens the area of investigation into memory, giving researchers an opportunity to devise more sophisticated methodologies for measuring memory processes, while allowing the necessary control over stimuli and their presentation conditions.

The purpose of the present research was to explore memory for pictures, to examine the limits in our capacity for remembering pictures and to study the constructive nature of encoding by manipulating temporal parameters and subjects' strategies. While an investigation of the encoding and retrieval of pictures would not be expected at this stage to lead to an entirely new formulation of memory, it can draw attention to discrepancies between findings of pictorial studies and assumptions of memory models which are applicable only to verbal studies. A recognition of these discrepancies might lead to more realistic theorizing.

In the next section, the past and present directions which have guided studies of pictorial memory are examined before there is a presentation of the present approach to the problem.

1.4 PAST AND PRESENT DIRECTIONS IN THE STUDY OF MEMORY FOR PICTURES

The historical movements which delayed the interest in the scientific study of cognition in psychology were also those which led to an emphasis on verbal material to the exclusion of other sorts of stimulus material. Most early memory experiments used lists of nonsense syllables [e.g., Ebbinghaus, 1885], words, letters or digits as stimuli. Following the trend towards behaviourism and learning theory which dominated psychology in the first half of this century, memory studies came to be identified as "verbal learning" studies. In these studies, phenomena important in other areas of learning, such as massed versus spaced practice effects, forgetting functions and proactive inhibition, were investigated. The rigid control over stimuli seen as necessary in these early experiments was readily provided by verbal materials.

The use of pictorial stimuli originated in perceptual experiments. During the 1920's the interest in Gestalt principles of perception stimulated many pictorial studies. In those days there was not always a clearcut distinction between perceptual experiments and memory experiments, and many of the "perceptual" studies required subjects to recognize or recall pictures from memory. During the early 1930's, a small group of researchers investigated memory for pictures. Influenced largely by European trends in psychology in which behaviourism played a minor role, researchers like J. Gibson [1929], Wulf [1922] and Bartlett [1932] studied subjects' reproductions of drawings they had seen earlier. Carmichael, Hogan and Walter [1932] received attention for their study of distortions in reproductions of shapes when subjects were given verbal labels with the shapes during presentation. The interest of perceptual psychologists may have encouraged these early studies of memory for pictures. Certainly, they were not cited by the verbal learning researchers studying memory.

During the 1950's there was a new movement in research, originating in the engineering field, which influenced verbal learning studies of memory. Information theory was used to define the meaningfulness of a stimulus in terms of binary digits of information [see Attneave, 1959]. For example, in an early study by Attneave [1955], dot patterns were varied in information content to show the importance of information theory for memory. In this study Attneave was using information concepts to control the meaningfulness of a pictorial stimulus, while he criticized attempts by perceptionists to predict the memorability of a stimulus on the basis of Gestalt notions, which are more difficult to define operationally. Information theory probably led to the merging of pictorial and verbal studies in the memory field, since dot patterns and shapes as well as verbal lists became acceptable stimuli to vary in information content for use in experiments. Experiments with random shapes have been common in testing information theory as well as learning theory. These pictorial stimuli are made by joining points randomly dispersed in a matrix into polygons [e.g., Vanderplas and Garvin, 1959]. The association value, or ease of naming shapes, has been measured to determine whether it affects their memorability. Association value has been used as a quantitative measure of meaningfulness; shapes can be more or less likely to resemble an actual object just as phrases of several words can approximate English at different levels [Miller and Selfridge, 1950].

The growing importance of information theory probably encouraged the retreat of learning theory from its dominant position in the memory field. There was a growing interest in the study of cognitive psychology at this time, concerned with thought processes rather than descriptions of relationships between stimuli and responses. Broadbent's [1958] model of an information-processing system in thinking hypothesized that a series of steps could be distinguished in the processes of attending to and remembering stimuli over time. This was the start of the growing area of research into attention, encoding, storage, retrieval and response mechanisms involved in memory. Hebb's [1960] presidential address to the American Psychological Association gave further encouragement to research into cognitive processes and theoretical arguments about memory.

There remain two distinct schools of research in remembering. Tulving and Madigan [1970] pointed out the difference between studies of "verbal learning" and those of "memory". Traditional verbal learning studies have continued to use experimental paradigms common in early learning studies, such as paired-associate learning and serial recall. On the other hand, cognitive psychology has used many paradigms including recognition and free and cued recall. These latter studies have also provided the opportunity for pictorial stimuli to be used to expand the narrow range of materials used in earlier memory studies. This does not imply that verbal learning studies have never used pictorial stimuli. The many studies comparing memory for words and pictures, best represented by the work of Paivio, signalled a broadening of the approaches in verbal learning during the 1960's. However, there is still little overlap between memory studies which are based on verbal learning and those based on cognitive or information-processing theories, and this prevents an integrative theoretical account of memory processes.

Present approaches to the study of memory for pictures can be divided into three main categories:

(1) those that study verbal recall of pictures, to determine whether or not pictures are simply translated into a verbal memory (the verbal loop hypothesis);

(2) those that investigate the role of verbalizing in memory for pictures, assuming that there is both a verbal and a pictorial memory code (the dual encoding hypothesis); and

(3) those investigating the properties of memory for pictures *per se*.

The first two of these areas will be discussed in the remainder of this chapter. The following chapter will explore the third.

1.4.1 The Verbal Loop Hypothesis

One extreme theoretical orientation concerning memory for pictures suggests that there is only verbal memory, and that pictures are encoded as words. This has come about partly from findings that namable pictures are easier to remember than pictures difficult to name. For example, Brown and Lenneberg [1954] showed that colours more easily labelled by subjects are easier to remember. These studies have led to an hypothesis implicit in much picture research that pictures are actually remembered by their labels. There is a large number of studies which measure recall of drawings by asking subjects to respond with a one-word label for each picture. When these data are found to replicate data from similar studies with words, it is assumed that both pictures and words have been remembered in the same way: with words [e.g., Glanzer and Cunitz, 1966]. This hypothesis implicit in much research was stated explicitly by Glanzer and Clark [1963], who first formulated this hypothesis to explain the way in which subjects can remember arrays of common black and white geometric shapes. According to Glanzer and Clark [1963, p.295], the subject "puts the information through a verbal loop:

(1) He translates the visual information into a series of words.

(2) He holds the verbalization and makes his final response on the basis of that."

In a later experiment, Glanzer and Clark [1964] claimed support for the hypothesis from the finding that the length of a subject's verbalization about a geometric figure correlates highly with the difficulty subjects have in drawing it and with high ratings of figure complexity. In the same vein, researchers have drawn evidence for the verbal loop hypothesis from studies of "communication accuracy", which suggest that recognition is not aided by lengthy verbalizations *per se*. Rather, the important variable is quality of verbalization, which can be measured by the extent to which a subject who has never seen the picture can choose it from an array of pictures on the basis of the recorded verbalization of a person who has seen the picture [Lantz and Stefflre, 1964; Smith and Larson, 1970]. Logically, this means that if a person describes a picture verbally to another person who has never seen it, both will have encoded the same information about it. This is an implausible hypothesis when applied to complex visual stimuli, though it might be useful in order to describe how visually presented words or stereotyped geometric forms are remembered. For example, a "black square" can easily be reproduced by anyone, just as a word can be written down after someone has uttered it. Complex pictures might conceivably be described in a series of paragraphs, but this would take several minutes, and it has been shown that memory for pictures is quite good when only short presentation times of a few seconds are used [e.g., Goldstein and Chance, 1970]. Pictures can be remembered when they are presented for less time than the half-second required [see Paivio and Csapo, 1969] to form a verbal label to describe them [e.g., Franken and Davis, 1975; Potter and Levy, 1969], although longer presentation times do aid recognition.

In this situation it is impossible to conclude that the pictures are remembered only with a verbal description. The verbal loop hypothesis also implies that people would have no memory for shapes of very low verbal codability, but in fact subjects can remember shapes which are difficult to label [Cermak, 1971].

Some studies have directly refuted the verbal loop hypothesis. Using a difficult recognition task which required subjects to remember detailed information about forms or photographs, two studies showed that length of verbalization about the pictures did not correlate significantly with recognition accuracy [Goldstein and Chance, 1970; Sang and Ross, 1970]. In a convincing refutation of the communication accuracy data, Murphy [1973] showed subjects successive glimpses of segments of a geometric pattern or successive verbal descriptions of the segments. Subjects given pictures recognized the whole pattern (presented visually) more quickly and accurately than subjects given verbal descriptions. If the picture segments were remembered as verbal descriptions, this difference would not have been found. (The verbalization needed to perform the task correctly was unambiguous.) This is similar to evidence for a purely visual code from Posner's work on physical and name matches of letters, showing that physical recognition matches do not require naming of the stimuli. On the basis of all the evidence, it seems implausible to suggest that pictures are remembered by words alone.

The verbal loop hypothesis may have played a role in the infrequent use of pictures as a memory stimulus. After all, there is no reason to study different kinds of stimuli if they are all remembered the same way, i.e., by language encoding. The verbal loop hypothesis has ignored evidence that there may be nonverbal encoding important in memory for pictures [e.g., Cohen and Granstrom, 1968; Posner, 1969]. The visual configuration of lines, which makes the pattern in a picture, may be remembered as an abstract structural description without the help of any language. Although words cannot account for all our memory for pictures, they do seem to be important, since pictures which are given labels, for example by the experimenter, are remembered differently from those not given labels. This literature does not assume that memory for pictures is based entirely on words.

1.4.2 The Effects of Verbal Encoding on Memory for Pictures

There is a long history of studies concerned with the effects of verbalizing on memory for pictures. Some researchers claim that labels can lead to distorted memories, while others show that labels increase memory for pictures. The main issue seems to be the appropriateness of labels given to pictures in the context of a particular task. The effects of verbalizing may be due to any of three factors:

(1) Dual encoding is important in memory for pictures because labelling a picture ensures that it is registered in memory in the verbal mode as well as in the pictorial or imaginal mode;

(2) Verbalizing about a picture forces the subject to encode the information in an organized manner easy to retrieve; or

(3) The greater labellability of a picture (or any stimulus) is a manifestation of its meaningfulness, and stimuli which are meaningful are better remembered than those that are not.

It is difficult to distinguish between these opposing theoretical positions on the basis of the vast literature that has been generated on labelling.

Distortion in the Direction of a Label

Some of the earliest studies of memory showed that reproductions of pictures from memory (i.e., drawings of remembered pictures) were distorted towards stereotypes. For example, Gibson [1929] suggested that many subjects reproduced a stimulus shape distorted in the direction of a verbal label which it may have evoked on presentation. In a study with a range of shapes, each labelled with an appropriate common noun, Bartlett [1932] also found that labels given seemed to affect later reproductions, so that many inaccurate reproductions resembled conventional representations of the labels.

The distorting effect of labels in recall of pictures was first systematically investigated by Carmichael, Hogan and Walter [1932]. These investigators presented subjects with ambiguous shapes and at the same time presented one of two plausible labels for each of the shapes. In immediate recall by drawing, subjects produced drawings more representative of objects resembling the labels given than the actual drawings shown; memory for the ambiguous pictures seemed to have been distorted in the direction of common pictorial representations of the labels given. Accuracy of the reproductions was judged and rated on a five-point accuracy scale by two of the authors. 26% of cases were classified as "most dissimilar" to the originals, and of these approximately 75% showed distortion in the direction of the verbal labels. A control group given no labels for the figures produced 45% of pictures distorted in the direction of one of the labels presented. Unfortunately, Carmichael *et al.* did not report the percentage of cases over all drawings that were distorted in the direction of the labels given. Although the study is suggestive, there may have been an inherent tendency to distort the pictures even when labels were not given.

Hanawalt and Demarest [1939] took issue with Carmichael *et al.*'s results and tried to determine whether reproduction distortions were caused by perceptual, memory or production difficulties. Using the same figures and general method of Carmichael *et al.*, Hanawalt and Demarest did not use the two lists of possible verbal labels in the same way. Instead, the ambiguous pictures were presented without verbal cues and later subjects were asked to reproduce these figures when the experimenter gave a label cue. For example, the experimenter asked for memory of the figure resembling a Roman 10 ("X") by saying, "draw the figure which resembled an hour glass" or, "draw the figure which resembled a folding camp stool". Subjects were asked to note, with each drawing, any verbalizations which had been associated with the figure when it was presented, and whether they really remembered the pictures they had drawn or whether they had attempted to represent the verbal label given by the experimenter. A control group received no verbal labels and reproduced the pictures in a free recall task. It is interesting to note that over 90% of subjects thought the figures resembled some object. There was a tendency for subjects to reproduce items in the direction of these "learning associations". Reproductions were classified by two judges as similar only to the original figures or influenced by verbal labels. For the control group, there was an increasing tendency to distort reproductions in the direction of verbal labels with delays up to one week, thus demonstrating the relevance of Carmichael *et al.*'s findings in both the immediate and delayed testing situation. Over time there was a tendency to be more influenced by a verbal suggestion before recall, and less by unexplained distortions not related to the labels given. Hanawalt and Demarest suggested that verbal association in either the learning or the recall phase might lead to distorted reproductions.

Bruner, Busiek and Minturn [1952] also improved Carmichael *et al.*'s design by anchoring the seven-point scale judges used to rate the accuracy of the drawings. A different drawing was used to anchor ratings of perfect accuracy for each subject; the standard of accuracy used was each subject's copy of the original shape. This helped to overcome the problem of individual differences in drawing ability which might obscure the findings. At the other end of the scale, the range of distortion towards a verbal label was anchored by each subject's drawing of the ambiguous figure placed in front of him while he was told to draw it as much like a particular critical label as possible. Bruner *et al.* replicated the Carmichael *et al.* study with shorter exposure times. The original design was improved by the addition of a control group given both labels for each ambiguous picture; control subjects distorted drawings towards one label or the other more often than subjects given no labels. However, some figures were more likely to evoke a particular label than others, and this complicated any interpretation of the findings. Bruner *et al.* suggested that distortion could occur if subjects did not have adequate time to perceive the stimulus as well as when subjects had "an alteration of forces of organization in a memory trace".

A later study [Herman, Lawless and Marshall, 1957] also used the Carmichael *et al.* stimuli and experimental design, with improvements, and reported a greater incidence of verbally influenced reproductions. A control group of subjects who knew in advance that they would have to recall the shapes made fewer biased reproductions than standard "incidental" subjects unaware of the memory task. In comparison with control subjects given no labels, the subjects given words in this study reproduced significantly more language-biased drawings. Unfortunately the results were marred by the tendency for the control subjects to label the pictures spontaneously. Over 60% of control subjects thought of labels for the pictures like those given by the experimenter to the groups given labels. From the Herman *et al.* [1957] results, there does not appear to be any difference between the three groups tested in the tendency to make accurate reproductions; only 1% of the drawings were accurate. The stimuli seem very susceptible to distortion, and the experimental groups were not unique in their tendency to be influenced by the verbal labels. Better stimuli for this paradigm would be those which the control subjects are not likely to label, and which would test the efficacy of labels supplied by the experimenter which the subject would not be likely to discover by himself. Other experiments have improved the design in this way.

Ronald Cohen has carried out several experiments which demonstrate a distorting effect of verbalization on memory for pictures when memory is tested by reproduction. Cohen [1966] presented subjects with a circle having a 90° gap on the left side, amongst four other pictures. All pictures were presented individually with a label provided by the experimenter. Actually only memory for the incomplete circle was measured and subjects were given one of two plausible labels for it, "a clock set at five minutes to seven" or "a clock set at ten minutes to eight". In this way, the 90° gap could be likened to the time on a clock with a 122.5° gap or a 55° gap between the minute and hour hands. Subjects were asked to draw all the pictures seen afterwards. Cohen measured the mean size of the gap in the circles drawn and found that the "five to seven" group averaged a 102° gap, while the "ten to eight" group averaged a 66° gap. This was evidence that labels do influence reproductions of pictures from memory. However, these results differed from those of a control group not shown the stimulus figure who had to choose circles representing the two time labels from an array of 20 circles, varying from 24° to 138° gaps in 6° steps. Control subjects choosing "five to seven" circles averaged gaps larger by 16° than the memory group, and control subjects choosing "ten to eight" circles also chose larger gaps than the memory group (differing by 7°). This seems to imply that the label produces only an incomplete distortion in the drawings from memory. However, this could also be due to the difference between the recognition method used to test control subjects' accuracy and the reproduction method used to test experimental subjects' accuracy.

Cohen also suggested that subjects in the free recall condition might have overlooked the fact that the hour hand on a clock points directly to each number on the dial only on the hour. Subjects in a recognition task would be more likely to see the greater familiarity of the correctly-positioned hands than subjects given no clues in a recall task. This tendency, if true, would be further evidence that subjects are influenced by stereotyped labels in memory more than by their accurate visual memories of the stimulus. Cohen's study should have included a control group given labels only during the recall phase of the task. The Hanawalt and Demarest (1939) study suggested that labels can distort reproductions whether they are presented with the stimuli or afterwards, during recall. This may be evidence that the verbal distortion is a response bias of subjects who do not remember very much about the pictures and use the experimenter's labels to confabulate in recall.

There is some evidence that the verbal distortion effect occurs in recognition as well as recall of pictures. Cohen [1966] found that memory of the incompleted circle could be distorted in the direction of a verbal label denoting the time on a clock equally well if the retention measure used were recall or recognition. On the other hand, Prentice [1954] failed to replicate the findings of Carmichael *et al.* in a similar experiment employing recognition as the measure of memory. Prentice's recognition distractors were based to some extent on actual distortions made by subjects in reproduction in the direction of verbal labels given. Subjects in the experiment did not have accurate memories; there were half as many errors as correct choices in recognition. The distractors chosen were not consistently in the direction of any verbal label, but this may have been because memory in general was fairly poor. Evidence in favour of the verbal distortion effect in recognition comes from a study which used distractors of varying degrees of similarity to original pictures and distorted pictures. Daniel [1972] used four 24-sided shapes chosen to represent a dog, camel, cat and duck. In each stimulus set there was a prototype animal and 11 variations of it formed by cumulative geometric changes of the points in each prototype so that the eleventh variation was least like the prototype. In a recognition task, subjects were shown four shapes, actually the midpoints (sixth variations) of each stimulus continuum. In recognition testing the 11 variations of each stimulus prototype were shown. Analysis of recognition errors showed that subjects more often chose the variations most similar to the animal shape than the dissimilar variations. Unfortunately, Daniel did not include a control group given no labels, so it is not clear whether subjects would have exhibited the same tendency to distortion towards a verbal label if they were left to label the pictures themselves. The studies by Daniel [1972] and Cohen [1966] suggest that verbal distortion effects occur in both recognition and recall.

These studies investigating distortion in the direction of a verbal label have many problems. Because of the difficulties in measuring reproduction accuracy, it is not known whether reproductions are as biased when labels are presented at the time of recall or recognition as they are when labels are given during presentation of the picture. Also, the intrinsic labellability of the pictures studied has often been ignored, and this must be carefully controlled to determine which labels are being associated with the shapes in memory.

Another field of research has provided experimental evidence which completely opposes the verbal distortion findings. This is reviewed in the next section. Teaching subjects appropriate labels for shapes has been found to aid their recognition. These differing findings may be reconciled by supposing that memory can be distorted more when labels are inappropriate than when they are relevant to the shapes tested.

Beneficial Effects of Labelling on Memory for Pictures

In most memory situations, labelling has been found to help rather than hinder recognition. There has been great concern, especially in the developmental literature, with experiments in which the experimenter controls labelling by teaching subjects to name shapes before recognition. Ranken [1963] found that teaching subjects names suggested by the shapes given in recognition aided their memory. Ellis [1968] gave subjects preliminary practice in learning paired-associates between labels and random shapes. Labels given were of three kinds: meaningful and relevant (most common associates of the forms), meaningful and irrelevant (meaningful words judged by four subjects to be inappropriate to the shapes) or meaningless and irrelevant nonsense syllables. Labels aided shape recognition only when they were meaningful and relevant. Some of the shapes were high in association value and others low in association value, but recognition performance did not differ significantly between these. The labelling practice probably overcame the disadvantages of difficult-to-label shapes. Ellis and Homan [1968] found that giving subjects any sort of verbal paired-associate practice before label training aided their recognition. Presumably subjects in certain tasks need practice in learning to label the stimuli. Two further studies by Ellis and his colleagues have shown that shape recognition is better when printed labels are presented originally with the stimuli, than when no labels are shown [Daniel and Ellis, 1972; Ellis and Daniel, 1971].

Similar findings for random shapes have been reported by Santa and Ranken [1972] who found that subjects given labelling practice with shapes recognized more than subjects given the same amount of time to look at the shapes with discrimination practice. When the presentation list was lengthened, the difference between scores of the naming group and the non-naming group was augmented. This suggests that the naming group was better able to encode and rehearse the list of items. Also, by filling the delay interval with a backward counting task, the advantage of the named group was reduced considerably, suggesting that the naming groups were using verbal rehearsal in the delay interval. Santa [1975] has also shown that name training can aid recognition of parts of a stimulus, in a rather unusual recognition task in which subjects decided if recognition shapes were parts of whole stimuli that had been presented. Meaningful labels were found to be most facilitative in the task, while irrelevant labels and "paralog" labels (e.g., "balap" and "latuk") were also facilitating. Subjects who both named the shapes themselves in training and heard labels supplied by the experimenter were superior to those hearing the names only. It would be difficult to explain the advantages of paralogs in ordinary recognition memory. In one of Ranken's recognition experiments, recognition for the group given no labels was superior to the group given paralogs to learn to associate with each shape. One explanation might be that subjects spontaneously labelled the pictures in every group, and these labels were more important to performance than the experimenter-supplied labels.

In fact, in the experiment mentioned, subjects were asked about labels they had supplied themselves spontaneously to the shapes. Subjects in the non-naming condition averaged more labels than those in the paralog group, showing that in fact they had the use of more relevant labels than the paralog group. Because of the poor manipulation of the labelling variable, this experiment did not effectively test the influence of labels on recognition of shapes. Thus it provides no firm evidence that inappropriate labels aid shape recognition.

These studies suggest that having an experimenter supply labels for shapes is not the important factor in remembering them. What appears to be most beneficial to memory for pictures is the skill of labelling pictures appropriately. Often subjects do this spontaneously, especially if experimenter-supplied labels seem irrelevant. The study by Ellis and Homan [1968] demonstrates the importance of calling subjects' attention to the possibility of labelling, which may not be automatic when subjects are given random shapes. There is support for the notion that the subject's own spontaneous labelling is the most important verbal factor in memory for pictures from studies which show that shapes resembling objects are easier to remember than shapes which cannot be labelled at all by subjects. Clark [1965, 1968] found that easily labelled random shapes (high in association value) were easier to remember than shapes which were difficult to label (low in association value). On the other hand, Price [1968] found that shapes of higher association value led to more false alarms in recognition than shapes lower in association value. This suggests that subjects may be biased towards responding to shapes which look more familiar or "thing-like". Two later studies by Price and Slive [1970a, 1970b] did find that subjects remember more high association value shapes than low association value shapes when an auditory label was presented initially with each shape. This may have made subjects pay more attention to the shapes and their labelling. When irrelevant labels were given for these shapes, recognition was poorer than when appropriate labels, often given spontaneously by subjects, were used.

The verbal distortion effect may be a manifestation of the inappropriate-label effect. The studies of labelling and verbal distortion are probably not comparable, so it is difficult to integrate these different findings. The distortion studies probably require more perfect performance from subjects than the labelling studies. If one of Carmichael *et al.*'s stimuli had been included in Ranken's [1963] experiment, a distorted memory on the basis of the label given would probably indicate a fairly good memory. The distortion studies only demonstrate a certain amount of memory loss, not the total memory loss indicated when the subject chooses a completely different distractor in a recognition set consisting of several globally different random shapes.

Verbal processes have been studied in recognition of faces in much the same way that they have been studied in recognition of other pictures. However, the labels subjects have used to describe faces are much more elaborate than those associated with shapes and there is evidence that some kinds of elaborate verbalizing may aid recognition memory. The verbal loop hypothesis suggests that pictures can be coded completely with words, but it is difficult to hypothesize that faces can be remembered by means of a list of adjectives. Frijda and Van de Geer [1961] found that subjects recognized faces best when there was high inter-subject agreement in three-word descriptions of the facial expressions portrayed [cf., Brown and Lenneberg, 1954]. All faces used were still photographs of different expressions of an actress. In this highly homogeneous set, verbalization would have to be limited to global statements about feeling or expression, since information about physical facial features (e.g., interocular distance, nose length) would be the same in all pictures. Perhaps verbalization was, in this case, an index of the meaningfulness of the faces: faces poorly recognized were those which were difficult to interpret meaningfully.

One study has shown that training in verbalizing faces aids the recognition of a new set of faces [Malpass, Lavigueur and Weldon, 1973]. On the other hand, Goldstein and Chance [1970] found no significant correlation between recognizability of faces and the frequency of verbal associations subjects made to them. Perhaps in certain cases verbalizing helps subjects to attend to significant features of the faces, while in normal circumstances subjects do not describe faces to themselves, and their verbalizations are not accurate enough to communicate to another person. In the Frijda and Van de Geer [1961] study and the Malpass *et al.* [1973] study, subjects' verbalizations could be used to aid their subsequent recognition, since subjects saw the labelled faces on the recognition task, whereas Goldstein and Chance measured verbal associations to their faces as a control of the meaningfulness factor, much in the same way as would a verbal study, by asking an independent group of subjects to associate to the faces. The memory task was given to a set of subjects who had no verbalization training or practice. It appears as though subjects in Goldstein and Chance's experiment verbalized in a different way, for different reasons, than subjects in the other experiments mentioned. Perhaps it is only the subject's spontaneous, idiosyncratic verbalization that aids memory.

But why should an appropriate verbalization aid memory for pictures? Dual encoding theory seems to provide a ready answer: encoding information from a verbal and pictorial mode at the same time in some way enriches the memory trace in a way that mere elaboration in one mode would not. In the specific case of memory for visual scenes, there may be an added benefit of verbal encoding because language is finite, logical and ordered, allowing for easier retrieval of information from memory than is allowed by the jumbled, infinitely complex memories of purely visual information. There is no evidence that memory for pictures consists entirely of words, but it is widely suggested that words provide a retrieval cue for pictorial information [e.g., Paivio, 1971]. To explain the effects of labelling on shapes, Santa [1975, p.294] suggested, "A name can provide the basis of an integrated representation and increase the retrievability of the visual information." This does not necessarily imply that having two "modes" in which information is stored is the important factor. The act of verbalizing may draw subjects' attention to salient features in a picture so that encoding is more organized, leading to better retrieval, than it would be if subjects were merely gazing at a picture without organizing it during presentation.

There is an alternative hypothesis about the effects of verbalizing on pictures, and that is that pictures easier to label are simply more meaningful to subjects. Thus verbalization is an index of meaningfulness, and it is not the words in memory *per se* that create the labelling effects. The confusion about meaningfulness is demonstrated by studies which define meaningfulness operationally as labellability [e.g., Noble, 1952].

Meaningfulness of a Stimulus

There is no doubt that meaningful material is easier to remember than meaningless material. One of the earliest studies to show that meaningful pictures are better remembered than meaningless pictures was Street's [1931] Gestalt Completion Test. Subjects were shown patterns of shapes which could be perceived as meaningful pictures. Street found that memory for these shapes was very high when subjects perceived the array as a meaningful picture. Recently, a similar study was carried out using ambiguous patterns which could be seen as faces [Wiseman and Neisser, 1975]. Recognition accuracy was greater when pictures were seen as faces than when seen as meaningless patterns, showing that the same stimulus could be remembered well or badly, depending upon the perceptual organization used by the subject. One study [Mandler and Day, in press] reported that familiar-looking polygons (e.g., those which resembled animals) were better remembered than similar polygons not resembling any concrete object. Wiseman and Neisser's finding that the same object can be perceived in different ways is positing meaning in the subject, rather than attempting to define it as an external variable. The ultimate meaningfulness of a stimulus must reside in the subject. But since it is difficult to predict how subjects will process a stimulus at any time, both general trends and individual differences must be studied.

Meaningfulness has been defined in many ways. "Labellability" has been used as an index of meaningfulness, as well as "familiarity". Familiarity has often been studied in word recognition and recall, and also appears to be important in memory for pictures. Familiarity has been studied in two ways. It can be studied simply as an aspect of meaningfulness, if varied exposures to an object lead to wider attributions about its meaning because it has been experienced in different contexts. On the other hand, familiarity has been defined as the frequency of exposure of an item, which may not always lead to greater attribution of meaning. Presenting a person with the word "shoe" five times in a list of items is probably not five times as meaningful as a single presentation of the word. Seeing the word "hybris" in two different contexts, on the other hand, might increase its meaningfulness considerably. Frequency of occurrence is an obvious operational definition of familiarity, but it may be limited in its usefulness.

Arnoult [1956] was concerned with the relationship between frequency and familiarity. He asked subjects to rate random shapes for their "intrinsic familiarity". He presented other subjects with these shapes in an incidental recognition task, and controlled the frequency of occurrence of the shapes in the list. The shapes could appear once, twice or up to 25 times in the list of 93 pictures. Recognition accuracy increased with greater frequency of the pictures in the list, reaching an asymptote with two or more repetitions. Differences in intrinsic familiarity apparently did not influence recognition, but memory subjects judged the more frequent pictures as more familiar in a post-recognition rating task. Arnoult did not control lag in the presentation lists, which is the number of items intervening between repetitions. This has been shown to affect judgments about the frequency with which pictures have occurred in a list which subjects are trying to remember [e.g., Hintzman and Rogers, 1973]. Despite this, there may be a consistent correlation between judgments of familiarity and actual item frequency in this type of task.

Two studies on memory for faces done by Yarmey contrast these two types of familiarity, one based on frequency of occurrence and another on a global criterion of meaningfulness. Yarmey [1971] investigated effects of familiarity on recognition in two recognition tasks, one using ordinary facial photographs as stimuli (unknown faces) and one using photographs of famous people. The familiar faces were recognized significantly better than the unfamiliar faces. Scapinello and Yarmey [1970] compared recognition of unfamiliar faces, dog faces and buildings when stimuli were given seven repeated exposures or only one presentation in order to control familiarity as a frequency effect. "Familiar" pictures were recognized better than pictures seen only once, suggesting that repetition leads to better encoding of a picture. Thus recognition accuracy was better with more familiar faces, using two different operational definitions of familiarity.

Recognition of faces of famous people has been compared with recognition of unknown faces in patients with unilateral cortical lesions [Warrington and James, 1967]. There was no significant correlation between results of the two tasks. This suggests that long-term familiarity may be due to cognitive processes different from those which lead to short-term familiarity in the course of an experiment. But the famous faces recognition task required verbal recall of the names of people photographed and this alone may have produced the processing differences. Much of the physiological literature suggests that speech processing is dominant in the left hemisphere, while face recognition may be dominant in the right hemisphere.

The studies of familiarity of pictorial stimuli do not support the conclusion from verbal studies that more unusual stimuli are better remembered than common stimuli. In verbal memory experiments, rare words, as defined by their frequency of usage in the English language, are easier to recognize than common words [e.g., Schulman, 1967]. Since the opposite finding is true in verbal recall [Deese, 1959] and since increased frequency of presentation aids both recognition and recall of words, various explanations of this effect have been put forward [e.g., Lockhart, Craik and Jacoby, in press; Paivio, 1971]. On the other hand, common pictures are easier to remember than unusual pictures. Complex photographs may be well recognized [e.g., Shepard, 1967] while randomly generated shapes seem to be poorly recognized [e.g., Cermak, 1971]. Thus the rare stimulus is poorer in recognition. Further evidence is provided in one study in which shapes, comparable in every way except meaningfulness and labellability, were tested in recognition. Mandler and Day [in press] compared recognition of random shapes with very similar shapes which could be easily interpreted as representations of concrete objects. Though the same type of recognition task was used for testing both kinds of shapes, meaningful shapes were remembered better than the random shapes. This is similar to the finding that high association value shapes in recognition are better remembered than low association value shapes. The more labellable and meaningful a shape, the better it is remembered.

Findings from picture studies do not necessarily contradict those from word studies. No one has suggested that nonsense syllables are better recognized than words, although the former are more novel stimuli. So it may not be familiarity *per se* which determines the memorability of a stimulus. Another investigation may provide a resolution of this problem. Standing [1973] found that pictures judged to be "vivid" in content are better remembered than "normal" pictures. Although Standing did not subject the differences to significance tests, normal pictures led to at least twice as many errors as vivid pictures in all recognition groups. This provides evidence that vividness is the important aspect of the well-remembered novel stimulus. Nonsense syllables do not seem particularly vivid, while words such as "travesty" and "sloop" may seem quite vivid to some subjects. A viable explanation of the recognition-recall difference for words is Lockhart *et al.*'s [in press] hypothesis that rare words are encoded better than words which are immediately familiar because they are subjected to more encoding of their constituent letters, phonemes and possible meanings. This leads to excellent performance in recognition, since the rare words are accessed by clues in the task, but in recall the rich trace is difficult to retrieve since it is not easily associated with cues present to subjects in that situation. It may be added that it is the vividness of rare words which helps their elaborated encoding. This could be tested easily in a verbal study, where it would be predicted that familiar words rated as "vivid" would be better remembered than familiar words rated as "not vivid".

In summary, there seems to be no conclusive evidence that the positive effects of labelling and meaningfulness on memory for pictures involve essentially different processes. One exception to this might occur when subjects remember very little visual detail about a picture. In this case they may use their knowledge of conventional drawings of the thing partially remembered in order to produce a confabulated reproduction. Sometimes subjects may remember only a label brought to mind when a picture was seen; then the verbal effect is quite plausible. However, the positive effects of labelling on memory can be explained easily as a manifestation of meaningfulness. Unfortunately there is no way to distinguish between these two concepts, as there are no pictures which are meaningful and difficult to label. Subjects can always label pictures, even if they merely say to themselves, "it is a picture". A methodological problem which prevents a comparison of meaningfulness and labelling is that meaningfulness is difficult to define. It may be an interaction of many variables such as labellability, a feeling of familiarity, frequency with which the stimulus has been encountered and vividness. No operational definition of this concept seems particularly satisfactory.

The central point in this discussion is that it is not parsimonious to discuss labelling effects solely as a manifestation of the verbal system. The fact that memory capacity is increased by labelling, as well as by other variables which can be seen as manifestations of that nebulous concept of meaningfulness, suggests that there is a theoretical construct of general importance which has not been fully examined.

CHAPTER TWO

MEASURING MEMORY FOR PICTURES

It is more difficult to measure memory for pictures (and scenes in the real world) than to measure memory for words, because in the former there is no unambiguous means by which the person can communicate visuo-spatial information to another person. Words are part of the communication system of language, but images and memory for pictorial detail are private, mental events which may have no automatic output mechanism. Thus, it is difficult to find a measure of memory for pictures which reveals the true nature of the subject's stored information about the stimulus. Several different measuring techniques have been used. Subjects are sometimes asked to describe what they have seen in words, but words cannot be used to describe all the configurational, spatial and aesthetic qualities of a picture. When subjects are asked to draw the picture they have seen, more spatial information and other difficult-to-label pictorial characteristics are revealed than is possible with verbal recall. However, drawing is a complex skill and subjects' production deficiencies make it impossible to separate poor memory from lack of motor skill in drawing. A technique combining drawings with attendant verbalizations is a recall method which overcomes some of the disadvantages of both methods, but may still lead to responses difficult to quantify to achieve a measure of accuracy.

Recently, many experiments measuring memory for pictures have presented subjects with cues, so that subjects are all equally advantaged by the testing conditions. For example, in cued recall or reproduction, subjects are given a part of a picture and have to reconstruct the rest around it by drawing or by manipulating blocks, sticks or puzzle pieces. This method minimizes the uncontrolled effects of individual differences in drawing ability or verbal fluency. It is not often used, however, because there are the same quantification problems in measuring subjects' accuracy of memory that are found in recall. One method which does overcome the quantification problem is the recognition paradigm, in which subjects are presented with a whole stimulus and have to make a judgment about whether it has been seen previously or not. Subjects do not have to provide any information about the stimulus, but they must communicate what they remember by discriminating the stimulus from alternative pictures.

The difficulty in choosing an appropriate methodology for measuring memory for pictures has been approached in many ways. Earlier this century, several researchers investigated memory for pictures by asking subjects to draw pictures they had seen after varying intervals of delay. These experiments produced rich descriptions of errors in memory and provided unique insights into the memory processes of their subjects. Because of their poor control over stimuli and their failure to produce quantitative measures of memory accuracy, these researchers were left behind in the mainstream of learning and memory experiments important in their day. Recently new experiments have attempted to revive the picture-drawing method of recall, but it has been almost totally replaced by the recognition method and by verbal recall. Each type of methodology has come to different conclusions about our memory for pictures, so there is an unavoidable suggestion that the different methods of measuring memory are not operational definitions of the same memory process. In this chapter, each different type of methodology for measuring memory for pictures is presented, followed by a discussion attempting to reconcile or at least integrate these differences in order to produce an appropriate methodology. An improved recognition methodology, the exclusion set paradigm, is described at the end of the chapter.

2.1 REPRODUCTION OF PICTURES BY DRAWING

2.1.1 Early Studies

A very early study investigating memory for shapes which used reproduction as a measuring technique was carried out in Germany by Wulf. In Wulf's experiment [summarized by Gibson, 1929], subjects were shown abstract figures and had to reproduce them from memory 30 seconds later. While drawing the figures, they gave a report of their experiences, describing images and meaningful ideas associated with each figure. (This is a potentially useful technique for distinguishing failures in memory from failures in drawing skill.) With longer delays between presentation of stimuli and testing of memory, Wulf noticed that reproductions showed an increasing tendency either towards "sharpening", exaggeration of some detail in the figure, or "levelling", omission or de-emphasis of a detail. Wulf thought sharpening or levelling could occur in three ways:

(1) *normalizing*: the drawing was distorted to resemble a familiar object;

(2) *emphasizing*: a detail or peculiarity of the figure was noted by subjects and exaggerated in reproduction;

(3) *structurally conditioned changes*: these occurred when the figure was reproduced as a more coherent figure than the original; for example, an asymmetrical figure was drawn as symmetrical.

A later study by J. Gibson [1929] was an attempt to extend Wulf's findings. Subjects were shown 14 straight-line figures or 14 curved-line figures one at a time, for 1.5 seconds each and then asked to reproduce as many as possible from memory. Each subject saw seven closed-line figures and seven open figures. Subjects were questioned about their drawings to determine the extent of their variation from the originals. One group of subjects saw the pictures only once, while another was shown the list of figures and then asked to reproduce them several times. Gibson wanted this latter group to have reached the criterion of perfect performance before delayed testing was begun, but this was found to be impossible. The main problem was that "no definite criterion could be found of having *learned* a figure, since there could be no arbitrary standard of the correctness of a reproduction" [Gibson, 1929, p.7]. This measurement problem in scoring reproductions has never been resolved. Apparently drawings of the fairly simple figures were never correct, and subjects developed particular habits in drawing which were distorting. To determine drawing adequacy, Gibson should have given his subjects practice in copying the pictures. Gibson analysed the 4,000 reproductions made in the experiment and classified six types of changes in the reproductions.

(1) *Object Assimilation*: Subjects distorted the drawings to resemble familiar objects. This was much like Wulf's "normalization".

(2) *Verbal Analysis*: Subjects altered their reproductions in the direction of a verbal memory, presumably a label the shape had evoked on presentation.

(3) *Figure Assimilation*: A blending or confusion of two figures occurred so that a reproduction of one modified the reproduction of another. This was the most common distortion.

(4) *Completion*: Sometimes the open figures were reproduced with gaps in the figure made smaller or else completely closed in.

(5) *Disintegration*: Some of the open figures were reproduced as separate parts, consisting of two or more scattered parts of the original.

(6) *Rectilinearity*: Curved figures were sometimes drawn as straight-line figures, although the straight-line shapes were never reproduced as curved. Gibson also found memory for straight-line figures somewhat superior to memory for curved-line figures, which could explain this effect, since curved figures may be encoded and retrieved inadequately.

Gibson distinguished between "object assimilation" and "verbal analysis". In the former, the figure is associated directly with a single familiar object or shape, while in the latter the reproduction "is wholly or partially a product of the verbal analysis rather than a representation of the form itself" [Gibson, 1929, p.19]. These linguistically-based representations were thought to contain less visual detail than object assimilation reproductions. Gibson also suggested that reproductions influenced by verbal analysis are perceived as parts due mainly to verbal elaboration, while drawings influenced by object assimilation are perceived as wholes and accompanied by more visual imagery. This was supported by questioning subjects about their experiences. This distinction seems to provide evidence against the hypothesis that verbalization effects are due only to the meaningfulness of a picture. But Gibson provided no quantitative evidence that some drawings contain only verbal analysis. The distinction is based on his own classification system of memory errors and feedback from his subjects. Hence it is a *post hoc* hypothesis rather than an experimental finding.

Gibson criticized Wulf's analysis on the basis of his drawings. He suggested that Wulf's drawings often consisted of two parts, which by their nature tended to be reproduced with either levelled or sharpened distortions. In fact, Gibson's figures are also narrow in range. The classification of distortions may depend on the stimuli used. A recurring problem in this research has been the difficulty of measuring distortion quantitatively.

Allport [1930] criticized the studies of Gibson and Wulf and made several methodological improvements. Instead of using a few adult subjects, he used 250 adolescents and tested memory for only two shapes, similar to those used in the previous experiments. He also introduced his own elaborate scoring method to attempt to make judgments of distortion more objective. He did not find evidence of either merging between two figures or disintegration, as Gibson [1929] found, but he did find the following common errors: size reduction, displacement of position, a tendency towards symmetry in asymmetrical figures, simplification, underemphasis of main features ("levelling") and an increase in jagged contours ("sharpening"). Allport's description of distortions in drawings relied less on subjective hypotheses about the reasons for common errors than did Gibson's or Wulf's descriptions.

A large study by Bartlett [1932] was the first to include drawings of people and other meaningful objects and to use subjects' oral recall of what they had seen as well as their reproductions. Two interesting errors were noted in his subjects' verbal recall of pictures of military men. In the study, 70% of Bartlett's subjects thought a detail in one face belonged to another face; this "transference" of detail showed that subjects were confused about the context of a detail they remembered. "Importation" errors, where subjects introduced details in recall which had not been presented, occurred for 65% of the subjects in immediate free recall. Bartlett also noted that the tendency to import detail may have increased over intervals of two weeks or longer.

In a study of "picture writing", Bartlett [1932] found the same errors in reproductions of shapes that had occurred in verbal recall of realistic pictures. Several other distortions were noted and added to the list of errors which he observed in subjects' recall of what they remembered. In this study, he trained subjects to associate 80 shapes (geometric "signs") with common words and asked them to draw these signs afterwards from memory when the word cue was given. Bartlett characterized distortions in the drawings by some of the errors listed below.

(1) The most common mistake in the reproductions was a figure reversal, where the direction in the picture was changed.

(2) Blending or confusion errors between pictures were demonstrated in two ways:

  (a) through transference, and

  (b) through invention of a new sign based on salient features of other signs.

(3) Bartlett also observed that in a meaningful picture, any item or detail detached from the central structure in space was liable to disappear, and "novel" detail was likely to be remembered more often than usual or unsurprising detail.

(4) Sometimes the signs were omitted completely. Presumably the word cue did not elicit any picture in memory.

(5) The words given seemed to affect the reproductions, so that the reproduced drawing might resemble a conventional representation of the label more than the sign itself.

(6) Substitution was a common reproduction error. Subjects sometimes drew another sign when given the word cue. This suggests that the signs were not seen by subjects as obvious representations of the words given, and subjects had to learn links between the signs and words in a paired-associate learning procedure.

(7) Occasionally (.003% of cases) a word evoked a completely new

sign from subjects. This importation error was again distinguished from transference errors, since in transference, detail was added to one picture when it had actually been presented as a detail in another picture; in importation new items were added.

The occurrence of the last four errors provides further evidence that inappropriate labels do not aid memory for shapes and that presenting words with pictures does not automatically lead to dual encoding of the picture.

2.1.2 A Further Examination of the Reproduction Method

Bartlett's study did not include an experiment in which realistic, meaningful pictures were tested by the method of reproduction by drawing. To determine whether subjects made the same sorts of reproduction errors with realistic pictures as with shapes, the author carried out a preliminary experiment to test the generality of the earlier findings. The experiment was intended to pretest items for one of the later experiments (Chapter Three). Instead of telling subjects beforehand about the memory task, subjects were asked to look at pictures, ostensibly in order to rate them for their aesthetic appeal. This task was intended to approximate "normal" viewing conditions for pictures where any memory for them is incidental. An intentional memory task encourages subjects to use special strategies for memorizing pictures, a task even less common in everyday life than it is for words. Seventeen complex, black and white line drawings, modified from pictures in children's books, were chosen for the stimulus pool. Four pictures were presented to each of 22 randomly-selected, naïve subjects, with the constraint that each of the 17 pictures was shown to five subjects. The pictures were drawn on 20 cm by 25 cm cards and presented successively to subjects in an individually randomized order. Subjects were given as long as they wished to make ratings on each picture's aesthetic appeal. A seven-point scale was used, ranging from "most appealing" (1) to "most unappealing" (7). Immediately after presentation of the pictures, subjects were asked to draw them from memory in any order. Because of the complexity of the drawings and subjects' poor drawing abilities, subjects were asked to describe what they were drawing in detail as they drew each picture remembered. Subjects' comments were recorded with a cassette tape recorder nearby.

The author and several colleagues classified errors in the drawings. No attempt was made to select a quantitative measure of subjects' accuracy of memory on the basis of the drawings, because subjects seemed unable to communicate their memories to their own satisfaction. A typical comment in the protocols was, "I know what it was, but I don't know how to describe it." Several errors reported in the earlier studies were noted again here. The types of errors classified are listed below.

(1) *Loss of Configurational Detail*: An accurate but simplified picture was drawn, so that only an outline or a few details were reproduced.

(2) *Omission*: An entire picture was forgotten and did not appear in reproduction.

(3) *Misplaced Items*: An item was remembered and reproduced, but in the wrong position relative to other items in the picture.

(4) *Interference between Pictures (Transference)*: One item presented in a picture was drawn out of context in a total scene representing another picture.

(5) *Importation*: Like "intrusions" in a word list, an item not shown to subjects in any picture was drawn in the context of a picture shown.

(6) *Semantically-based Confusions*: An item was drawn which, though unlike any original item in the picture shown, could be described by the same verbal label.

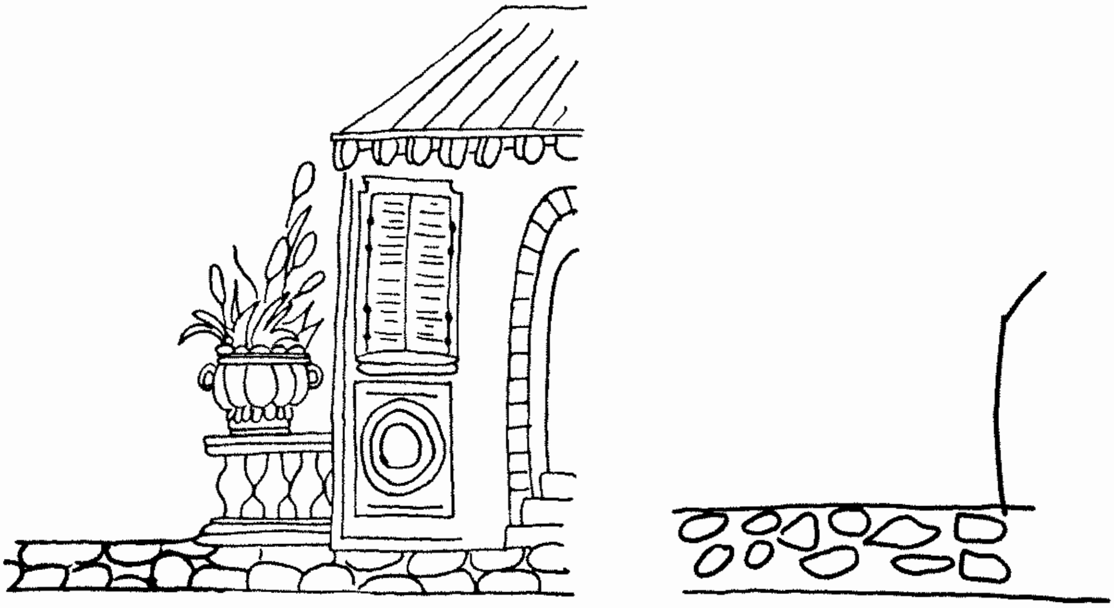
(7) *Confabulation (Structurally-based Confusions)*: Some accurate visual detail was present but it was used to draw an incorrect item.

All subjects showed loss of detail from the originals in their drawings, misplaced items from their correct position in the picture, and drew inappropriate items incorrectly. In other words, the drawings were not accurate reproductions. Only one subject forgot a picture and omitted it completely from his reproductions, however. Figure 1 shows examples of types of errors in reproduction alongside the corresponding stimuli presented to the subjects. In Figure 1(a) there is hardly any retention of detail; only the global outline of the path and house remain. This is an extreme case of loss of detail. In Figure 1(b) the clock, window, chair and mat drawn are all correct items, but they have no visual similarity to the original items shown. It is as though the subject has used pictorial synonyms for the real items; in other words, they are confusions of structural detail which are semantically correct. Misplacement of items is illustrated well in Figure 1(c), where all items are misplaced, including the duck which appears to have changed its status from a pattern on a blanket to the real thing. These three pictures do contain extreme errors, but all the drawings made contained these types of errors.

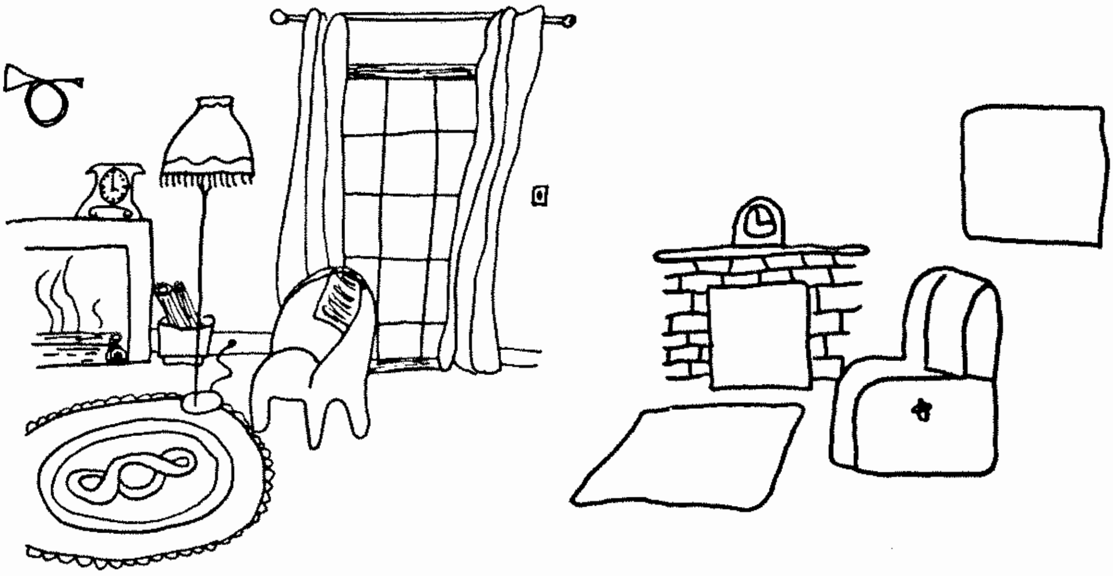
Interference between pictures, called transference by Bartlett [1932], can be seen in Figure 1(d). The lamp in the dining setting is an interpolation and may have been borrowed from the original in 1(b). Three subjects made obvious transference errors.

Importation errors were more common; 15 subjects had one or more reproductions with a completely new item. For example, in Figure 1(e), the sun in the window does not occur anywhere in the series. It is also interesting to note that this subject has separated the cabinet and sink in the kitchen, a severe misplacement of detail. The wavy line in the window seems to be another importation, this time of a bird, but it could also be confabulation on the lines in the calendar of the original.

One error was discovered in this experiment which has not been discussed by the previous researchers. Sometimes accurate detail in the pictures was reproduced in the drawing in a highly inappropriate context. It appeared as though subjects had retained disjointed remnants of information about pictures seen without encoding an entire pictorial event. When these pieces had to be retrieved and used in recalling a complete scene, subjects "confabulated" on the details remembered. Thus they had some structural but little semantic information about the picture. Eighteen subjects made confabulations in recall, and Figure 1 demonstrates several examples of this type of error. In Figure 1(f) the

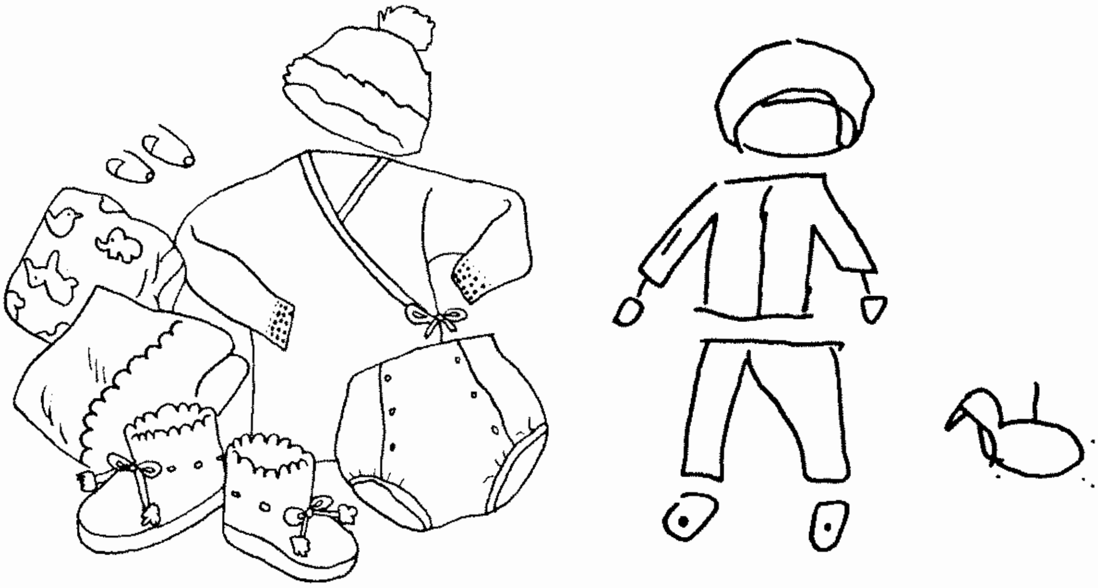


1a.

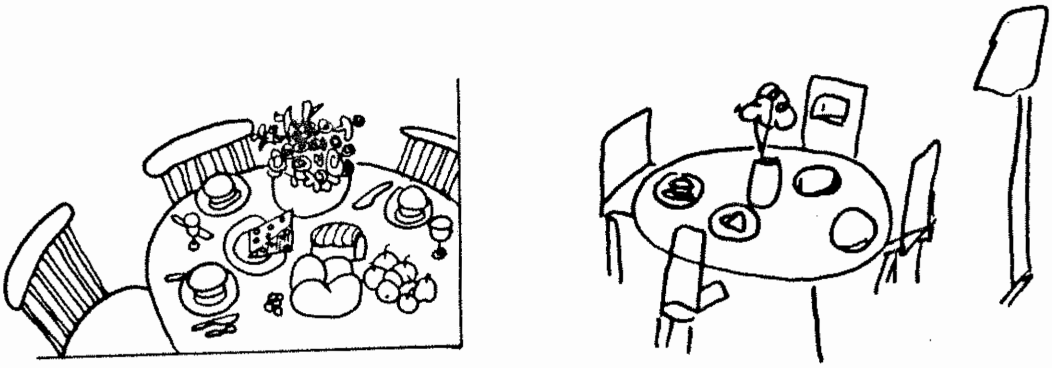


1b.

Figure 1. Examples of stimuli used in the preliminary study of reproduction with subjects' drawings from memory. Original pictures are presented on the left, and reproductions are presented on the right.

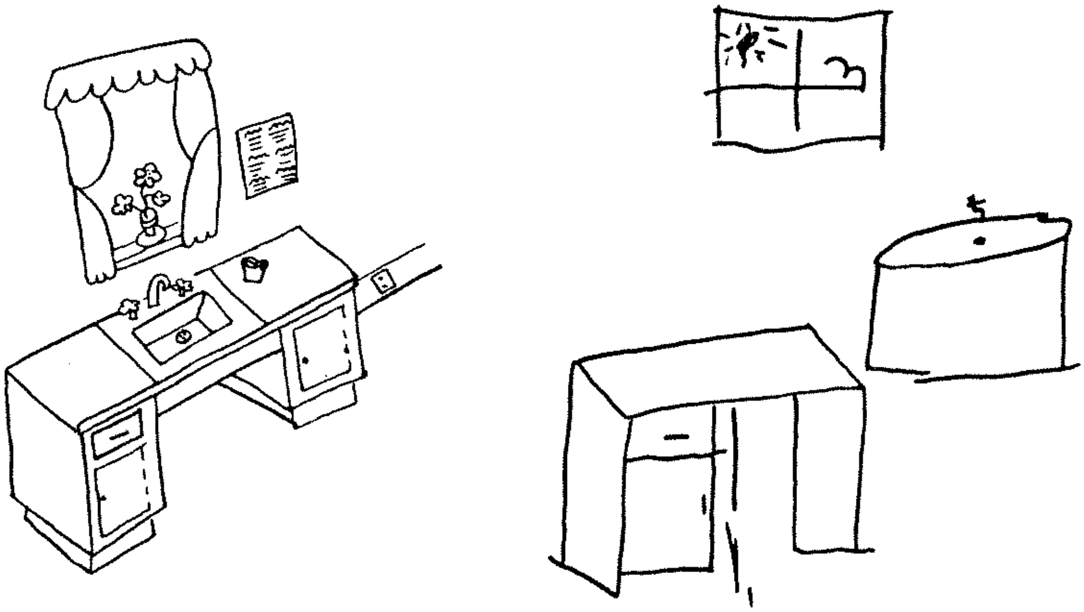


1c.

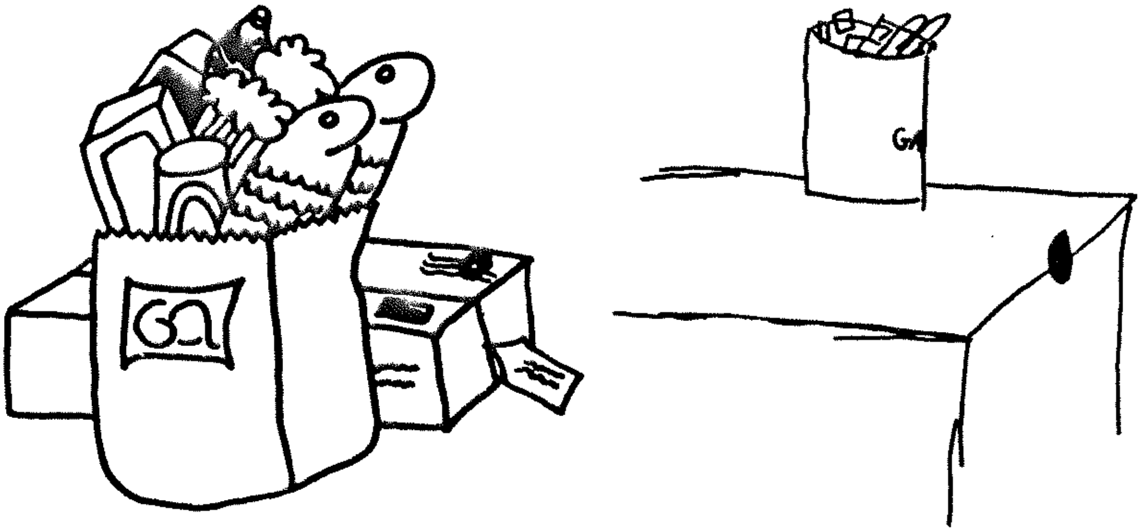


1d.

Figure 1. Examples of stimuli used in the preliminary study of reproduction with subjects' drawings from memory. Original pictures are presented on the left, and reproductions are presented on the right.

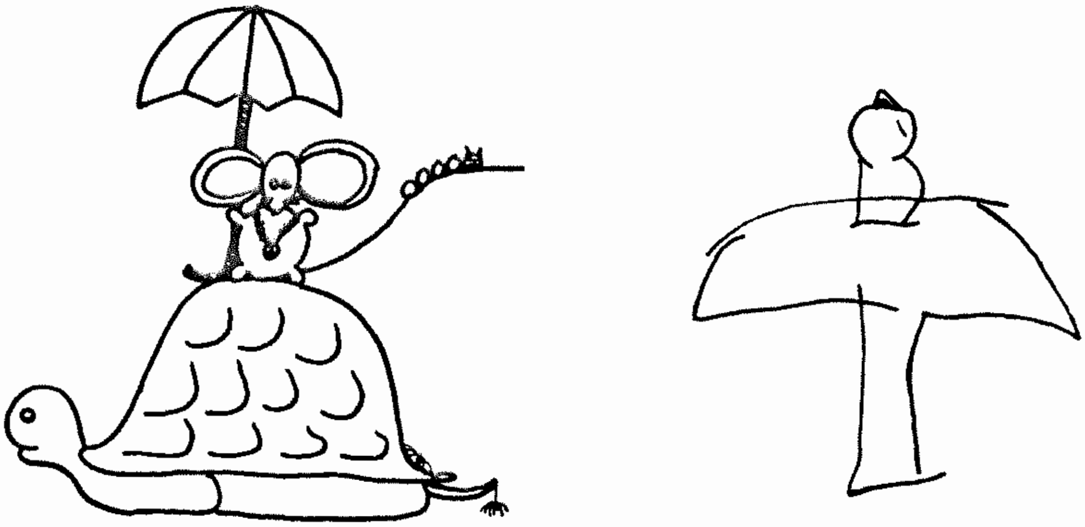


1e.



1f.

Figure 1. Examples of stimuli used in the preliminary study of reproduction with subjects' drawings from memory. Original pictures are presented on the left, and reproductions are presented on the right.



1g.

1h.

Figure 1. Examples of stimuli used in the preliminary study of reproduction with subjects' drawings from memory. Original pictures are presented on the left, and reproductions are presented on the right.

subject is attempting to recall the grocery bag, but according to his protocol the scene is "a rubbish bin with fish and other things ... fish were in that position ... I think it was on a wharf or something." It appears as though this subject had knowledge of fish and other items in a container, a rectangular object below and a confusing sign on the container such as "GA". This semantic information and visual detail about the picture was combined in a creative but incorrect reproduction which is a plausible picture.

The bizarre stimulus in Figure 1(g) was drawn by one subject as a "toadstool (spotted I think) with a little animal on it with a little thing on its head ... it may have had two ears but I got the impression it had a hook on its head." The hook-like ears of the mouse have apparently become a hook, while the toadstool may be the memory of the umbrella's shape or a combination of umbrella and turtle. All subjects who were shown this picture produced confabulated drawings from memory.

A final example of confabulation appears in Figure 1(h). It appears that the subject has remembered a monkey and an animal in a cage, and combined these into the drawing of the caged monkey, although this subject reported that she was unsure of the monkey's position. The adjacent rectangles in the cage have become a pile of books, apparently, since the cage is round and more similar to the back of the monkey's chair. Oddly enough, another subject drew a cage very similar to the chair as well. The subject was unsure of the clown in the picture and had to be persuaded to attempt to draw it. The unconnected visual detail remembered about the clock and "clock face" may have been confabulated upon to produce this picture.

Confabulation errors suggest two things:

(1) that subjects can remember parts of a picture and have no memory for the whole, and

(2) that subjects' communications about their memories are greatly influenced by their expectations about the requirements of the task. Confabulations seem to have occurred during retrieval, when subjects felt they had to draw a complete picture rather than only the odd fragments they actually remembered. It

> seems unlikely that subjects actually perceived a grossly distorted picture in the first place. Thus the distortion must have occurred in memory or during responding.

These are important points to consider in the study of memory for pictures. The hypothesis that pictures are not always encoded adequately is tested in a series of experiments later in this study.

The different types of errors catalogued by different researchers may be put together to form a comprehensive list of the distortions which may occur under different testing conditions. However, the difficulty of defining quantitatively the amount and type of distortion in a drawing limits the value of this research. It does not offer any information on the goodness of memory, or how much of each picture a subject has remembered. It only gives hints suggesting ways in which the memory process breaks down, so that the picture is not remembered completely.

2.1.3 Later Uses of the Method of Reproduction

A few other studies have attempted to measure goodness of memory for pictures using the method of reproduction by drawing. These studies have attempted to quantify accuracy of memory. One study provided information on decay of reproduction accuracy over time, without being concerned with the types of errors made by subjects. Rock and Engelstein [1959] asked different groups of subjects to reproduce from memory one shape which they did not realize they would be asked to recall, after various delay periods up to one month. Reproductions were rated on a 10-level scale of similarity to the original drawings by five judges, and were found to have decreased in their similarity to the originals with increasing delay. Rock and Engelstein reported a significant difference in median reproduction scores from 15 seconds' delay to three weeks' delay with a seven-sided polygon. As a control after testing, some subjects were shown the figure again and asked to copy it while it was present. Copies of the figure were more accurate than delayed reproductions, showing that distortions in drawings were not due merely to production difficulties. A curvilinear shape tested was more poorly reproduced than the polygonal shape at every delay period tested, suggesting that stimuli differ in the memorability. Rock and Engelstein's study is the only one so far discussed which has used a formal rating measure of a drawing's accuracy.

Some studies comparing memory for pictures and words have asked subjects to recall the pictures by drawing. In a typical experiment, subjects see a list of concrete nouns, half portrayed as words and half as conventionalized drawings, and recall more of the pictures than the words during testing [Kaplan, Kaplan and Sampson, 1968; Sampson, 1970]. Scoring of memory accuracy is based on the total number of items correctly recalled. Drawings are scored as either correct or incorrect, presumably by the experimenters.

An extension of this methodology was carried out in an experiment where subjects were shown either pictures or words and asked to recall them as they were presented or in the opposite mode [Fischler and Puff, 1971]. No differences in any of the four conditions was found. This finding is in opposition to Sampson's [1970] finding that pictures were recalled better than words under same-mode recall conditions. The Fischler and Puff finding stands in contrast to a large body of studies which demonstrate that pictures are easier to remember than words. This finding is probably due to the fact that subjects had only two minutes to recall all 16 items presented. This may reflect an overload on retrieval processes which have a limited capacity [see Mandler, 1975]. Neither Fischler and Puff nor Sampson and his colleagues report any difficulties in determining what each drawing is meant to represent. These studies would be more convincing if they addressed themselves to the problem of scoring reproductions for accuracy. For example, they should provide a report of inter-rater reliability in determining the number of items that have been recalled in the drawings.

The Kaplan and Sampson studies encountered an odd problem which suggests that encoding of pictorial detail was minimal in the task. Many subjects were biased towards representing items as pictures. 20% of subjects in the Kaplan *et al.* [1968] study and 26% of subjects in the group tested after one day's delay in Sampson's [1970] study showed reversals by recalling words with drawings or pictures with words. Most of the reversals were biased in the direction of representing words as pictures. Sampson [1970] suggested that the word reversals may have been due to a bias in the subject pool towards people with sharp imagery abilities. This implies that good imagers do not know the difference between their fanciful images and things they have seen. A more feasible explanation might be that subjects who are not trying to remember the stimulus list do not encode as much information about the mode of presentation of the stimuli as do intentional subjects. If pictures are more vivid and memorable than words, such subjects may vaguely remember that the list had many pictures in it. Thus, when they are able to recall a type of item from the list (e.g., an arrow), they may be biased towards recalling this in a stereotyped pictorial form rather than as a word. This may be seen as further evidence that pictorial qualities of letters and words are not usually noticed in perception. Otherwise, remembering a word would imply that subjects remember a picture of the word as they saw it printed originally. If pictures are more "arousing" in terms of galvanic skin response than words [Kaplan *et al.*, 1968] or arouse more attention than words, then it is not surprising that subjects remember more pictures than words and remember the list as having contained more pictures than it did. What is odd is that subjects do not remember the mode in which the item was presented. It throws doubt upon this recall method, because there is no guarantee that subjects remember anything about pictorial qualities of the pictures. They may be drawing conventional representations of the pictures when they remember only a word and some information about the typical mode of presentation. It would have been informative if subjects had been asked in these studies how many words and how many pictures they had seen. If the suggested hypothesis is correct, they should have been biased towards responding that there were more pictures than words.

The only study mentioned in this section which used a quantitative measure of subjects' accuracy of memory for a single picture was that of Rock and Engelstein [1959]. They successfully showed decay in memory for a shape over one month. Judges rated reproductions for their similarity to the originals. Measures of accuracy based on physical attributes of responses would provide a more reliable and consistent means of quantifying memory, but this is difficult to achieve if subjects' drawings do not reflect their memories as much as their drawing ability. The other recall methods [Fischler and Puff, 1971; Kaplan *et al.*, 1968; Sampson, 1970] measured only the subject's memory for a class of items which could be represented by a drawing. Since there was no reporting of the accuracy of individual drawings, the memory scores do not provide information about the content of memory for pictorial detail.

2.2 OTHER METHODS OF REPRODUCING PICTURES

Some recent studies of recall of pictures have attempted to control accuracy by limiting subjects' responses to a finite set of alternatives, so they can be quantified. Sheehan [1967] showed subjects pictures containing an array of coloured geometric objects, and asked them to reproduce their images of the pictures after the stimulus was removed, by arranging a pattern of wooden blocks corresponding to the geometric shapes seen. Accuracy was measured by counting the number of elements in each array not in the same position as in the stimulus pattern. Subjects had previously constructed the arrays while the stimulus was present, and these earlier reproductions provided standards with which subjects' memory reproductions could be compared. Sheehan also took an "inversion" score for each reproduction: the number of pairwise order inversions of the blocks needed to transform the memory reproduction into the perceptual copy. This method of reproduction seems to be a good one, since production problems should not greatly affect subjects' performance, and differences in reproductions could reflect real differences in memory accuracy. However, there are some problems with the measurement of accuracy used in this paradigm. Memory for parts of the pattern as well as the whole should be taken into account, and subjects given credit for remembering part of the pattern even if it is not reproduced in its correct position. Also, some inaccurate reproductions may indicate more serious losses from memory than others, even though the same number of incorrectly placed blocks occurs in each. For example, a pattern which includes a square could be reproduced incorrectly with a few extra blocks as either a rectangle or a U-shape. However, the latter might be a more serious deviation from an accurate memory reproduction than the former. These accuracy differences could be quantified, yet it may be difficult to take into account the many qualitatively-different types of error than can occur in order to produce a single accuracy score.

Another interesting recall method was used by Posner and Konick [1966] who tested memory for the position of a circle drawn at one of 12 positions on a line segment by asking subjects to draw in the position of the circle on a line segment the same size. Posner and Konick scored recall as the difference in absolute distance between the position of the circle on the stimulus line and the position drawn in recall. They did not report any measurement difficulties in determining the accuracy of memory for this specific type of information. Subjects would presumably have few production problems in performing the task, making the testing method an excellent one for measuring memory for finely discriminated spatial information.

These two methods have potential as a form of cued recall. Subjects' ability to reproduce the stimuli is not quite so influential in affecting memory scores when the reproduction task is easier and the responses which can be given are limited.

2.3 THE USE OF WORDS TO RECALL PICTURES

As has been discussed already, there is much literature on recall of pictures where the responses made are verbal. In free recall of pictures, a list of drawings, usually representations of concrete nouns, is shown to subjects and afterwards they are asked to recall a one-word label for each picture in any order. In this paradigm it is assumed that pictures evoke logogen-like labels [e.g., Morton, 1970] and that this is an important part of picture encoding. These verbal recall studies measure only one part of memory for pictures: memory for the class of objects to which the represented item belongs.

Most of the recall research using pictures has been concerned with showing that the labels denoting pictures are recalled better than visually-presented words. One study which did not find the usual superiority effect with pictures [Ternes and Yuille, 1972] is the only one to mention the problem that pictures are labelled differently by different subjects. This occurred despite the investigators' attempts to choose pictures which could be unambiguously labelled. Ternes and Yuille reported that "a liberal scoring method was employed" for the picture data. No doubt this problem has occurred in other verbal recall studies. It may not be wise to assume that all subjects will use the same generic terms in classifying pictures seen.

Some studies have used much richer verbal descriptions to determine how well a subject has remembered a picture. This method has already been mentioned in discussing Bartlett's [1932] work and was used in the preliminary study reported above. Standing [1973] showed subjects lists of pictures and asked them to recall the pictures with a verbal description instead of a one-word label. Standing instructed subjects to make clear descriptions, so that judges could identify the pictures to which the descriptions referred. For scoring the number of pictures correctly recalled, five naïve subjects were asked to rate the descriptions as good, mediocre or erroneous descriptions of the actual stimuli. The pictures were complex photographs which were more difficult to label than the simple line drawings used in other studies. Descriptions given were generally six words in length; this may have been due to the time limit of 40 minutes for recall of the 200 photographs. Subjects averaged 11 "erroneous" descriptions in recall. There were more extra-list intrusions in this test than in a similar test of 200 words given to another group of subjects. The rating procedure was subject to another error that does not occur in word recall: Standing estimated a mean probability of 0.13 that a subject's description could fit a picture by chance. This was based on a control test attempting to fit the actual descriptions given to a completely new set of 200 pictures. Of course, it would be possible for subjects recalling word lists to respond with correct words by chance, but this is easily taken into account in scoring by using a correction based on the responding of a control group asked to write down a list of words. However, the presence of erroneous descriptions was not as serious a problem as in the recall of the pictures.

Another verbal recall technique used with pictures has been to probe for pictorial details by asking questions. Bartlett's [1932] early study employing this technique has its modern analogues. For example, Smith and Nielsen [1970] showed subjects caricatures of faces one at a time and measured recall by asking subjects if probed features (eyes, nose, ears, eyebrows or mouth) were large, medium or small. Subjects were aware of the features to be probed before presentation. Thus subjects had to translate each picture into a verbal response on the basis of the size discrimination task. This was a rather specialized task where subjects had to attend to and remember little about each picture. In an experiment using more information from subjects, Kosslyn [1973] explored memory for detailed line drawings, all of them with namable features. Subjects were shown 10 pictures and afterwards recalled the pictures by saying "true" or "false" when the experimenter recited the name of details in the pictures, which could be correct or incorrect. Subjects responded "true" only to details actually seen. This method is a potentially powerful one for measuring the amount of visual detail which can be recalled. In both experiments, the verbal response is a forced choice from a limited set of alternatives. These verbal reports can easily be scored for correctness or incorrectness, but they may force subjects to encode information about the pictures in a particular way without tapping subjects' full knowledge about visual detail. Apparently, not all subjects encode pictures verbally in the same way; some subjects may rely more heavily on verbalized encodings than others [Kosslyn, 1973].

Some studies using the verbal recall method with pictures have attempted to measure nonverbal information. Frost [1971, 1972] has examined clustering in recall of lists of simple pictures, using a paradigm previously used with words. She found evidence of clustering on the basis of shape or orientation similarities between pictures, which is not found with words. This provides evidence that configurational characteristics are more important in encoding of pictures than words. This replicated findings of an earlier study [Bousfield, Berkowitz and Whitmarsh, 1959] in which clustering was demonstrated with pictures when no verbal responses were required which could mediate the clustering. In this case the reproduction method gave more conclusive results about visual clustering than the verbal method.

Verbal recall methods of measuring memory for pictures enable better quantification of memory accuracy than drawing methods of reproduction. They are not absolute measures of accuracy, but they can be employed in experiments to compare the relative accuracy of memory under different experimental conditions. Verbal recall can only give information about those aspects of memory which can be communicated in words, which may not include many important aspects of the remembered picture.

2.4 RECOGNIZING PICTURES

Most studies measuring how well pictures can be remembered have used the recognition paradigm. In recognition, subjects always have a stimulus to examine and merely indicate whether it is "old" (i.e., one they remember) or "new". In recall of any kind, subjects have to make a more complex response and communicate to the experimenter some of the information about the stimuli which they have in memory. The cues which aid retrieval of information from memory are different in recognition and recall. In recognition, the subject has an entire stimulus to respond to, and if the stimulus is actually the original memory item, the subject has all the relevant cues about the object itself (but perhaps not the context in which it was seen) needed to make a decision in recognition. On the other hand, in recall the subject has only situational cues and any aids given by the experimenter during testing, to aid retrieval of the information. Subjects in a recall experiment have no other physical cues in the environment which might remind them of their perception of the stimulus in the previous situation.

Because recognition of words is often found to be higher than their recall [McDougall, 1904, cited by Kintsch, 1970], recognition tests may be able to access stored information which is unavailable in recall. This suggests that the recognition process probes more deeply into memory than recall. Also, in studying memory for pictures, the recognition method has the potential to tap memory for nonverbal information which verbal recall cannot access. For these reasons, recognition is the method of choice for most experiments investigating the extent of the subjects' knowledge about a picture. Since it does not depend on any complex production skill, recognition is also superior to the method of reproduction by drawing. However, there are certain problems with the recognition method. Subjects are usually given several stimuli, from which they must choose the old, remembered items. This creates the problem that subjects may guess the correct item by chance. It is difficult to use a constant statistical correction for guessing in this situation, as there are individual differences in subjects' strategies. Some subjects guess readily when they do not remember anything, while others are extremely cautious responders. Also, subjects may use partial information in different ways to maximize accuracy scores or minimize incorrect responses. (Subjects may consider one of these goals more important than the other.)

The findings of studies which have measured recognition memory for pictures may be divided into two classes:

(1) those which estimate a very high capacity for memory, and

(2) those which estimate that memory capacity is much lower.

Even when the same stimulus is used, estimates of subjects' recognition capacity vary with several factors in the recognition procedure. For this reason, the percentage estimates of recognition capacity cited by researchers in this area must not be taken at face value as demonstrating an absolute amount of memory. Accuracy estimates may be used to compare experimental groups to investigate the factors which aid or hinder memory, but comparisons across studies may not be valid. In the discussion to follow, the difficulty of a recognition task is attributed largely to the similarity between items in the test set.

2.4.1 Studies Demonstrating High Recognition Capacity

Several studies have shown that subjects can remember hundreds of pictures presented sequentially for brief exposures, if memory is tested using a two-alternative forced-choice test (where each old item is paired with a completely new distractor picture). These findings have been cited as evidence that memory for pictures is quasi-photographic and of immense capacity. A study by Nickerson [1965] was the main starting point for this work. In his experiment, subjects were shown 200 black and white photographs for five seconds each. Then a list of 400 pictures was presented immediately after, of which half were duplicates of the stimuli and half were "new" (i.e., distractor) items. Subjects responded to each of these by saying whether each picture was "old" or "new", and on average recognized 95% of the pictures correctly. Many subjects achieved perfect scores. The delay between an item's first presentation and its second occurrence in testing varied from about three minutes to 16 minutes. Over this delay period, recognition dropped from about 98% to 87%, which was a significant difference.

Nickerson [1968] then retested different groups of these earlier subjects after delay periods of one, seven, 28 or approximately 360 days. Using the same procedure as in the earlier study [Nickerson, 1965], old test items were interspersed in a recognition test with 100 new items. Of the 100 old items, half had been used as distractors in the earlier experiment. The probability of a correct response dropped from 92% with one day's delay to 63% with a year's delay. Pictures seen twice in the first experiment, that is, original old items, were better remembered than pictures which were remembered "incidentally". (These latter pictures were distractors used in the first experiment, and it was assumed that the subjects would have had no need to attend to them in detail if they could make an immediate rejection of them as "new" pictures.) Nevertheless, after one day, about 78% of these latter pictures were remembered. This has been cited as evidence that pictures can be encoded automatically, since a picture can be quite well remembered even if it is seen once and the subject is not trying to remember it.

A widely-quoted study was carried out by Shepard [1967]. He showed subjects 612 coloured photographs taken from magazines. After presentation of the items, subjects were shown 68 pairs of items each containing one correct old item and one distractor item. Subjects had to choose the old item in each pair. The median level of correct performance was 98.5%, again evidence of a very high memory capacity. With delayed testing, subjects increased their recognition accuracy from 96.7% with no delay to 99.7% with two hours' delay, but after 120 days recognition was only 58% correct.

This same memory paradigm was used in two later studies which constituted a more ambitious attempt to test memory capacity. Standing, Conezio and Haber [1970] showed five subjects 2,560 photographs for 10 seconds each over two to four days and then tested them immediately for recognition with forced-choice pairs. Subjects responded correctly to 90% of the pairs. Standing *et al.*'s work was only suggestive, since the size of experimental groups ranged from two to four subjects. However, their results were fairly consistent and subjects had high memory capacity for pictures presented for one second, two seconds or 10 seconds, even after 30 minutes' delay or when subjects were shown pictures four days before they were tested.

Standing [1973] has perhaps offered the last word in these capacity studies by presenting subjects with 10,000 pictures for five seconds each over five days and testing them with 160 forced-choice pairs in recognition. Subjects recognized 99.7% of the pairs correctly.

The high capacity of memory has not only been demonstrated with photographs. One study compared memory capacity for three different kinds of pictures with memory for short verbal descriptions [Nelson, Metzler and Reed, 1974]. Different groups of subjects saw either photographs, embellished line drawings, unembellished line drawings or verbal descriptions of the same scenes. Pictures were recognized better than words when stimuli were tested in forced-choice pairs in which an old stimulus was presented with a completely new stimulus of the same type. There was extremely high recognition of all types of pictures from about 98% in immediate testing to over 85% after one week's delay. The three types of pictures did not differ in memorability and the authors suggested therefore that the high capacity of memory for pictures is not simply due to the extra configurational details which they contain in comparison with words.

Recognition capacity for pictures has also been studied developmentally. Using a paradigm similar to Shepard's [1967], Nelson [1971] found that subjects in grades one, four and seven remembered about 69% of 36 pictures presented, although groups given an 0.8 second presentation rate for each picture remembered significantly less than groups given a five second presentation rate. His stimuli were paintings rather than photographs of objects. A study using younger children [Brown and Scott, 1971] found that three- to five-year-olds recognized 98% of 44 drawings of objects. Recognition was tested immediately and again after either one, two, seven or 28 days with the same pictures and with new distractors. Over the one month period, recognition declined to 78%. Brown and Scott's [1971] findings replicate closely those of adult studies, while Nelson's older subjects remembered somewhat less. Nelson used forced-choice pairs in recognition testing, while Brown and Scott used a "continuous recognition" paradigm [Shepard and Teghtsoonian, 1961]. In the latter procedure, presentation items are not separated from test items. Instead, a list of single items is presented, some items being repeated, and subjects are asked to make an "old" or "new" judgment about each item. This is similar to the paradigm used by Nickerson [1965] in his recognition study. Nickerson's capacity estimates were only slightly lower than the estimates of the other researchers using forced-choice pairs in testing.

The very high estimates of memory capacity in these studies should not be taken at face value. In the recognition paradigm, subjects can guess which picture is the correct test item. The influence of guessing on a recognition test is determined largely by the number of alternatives with which the subject is provided. For this reason, guessing corrections are often performed on the data to render more meaningful scores. For example, in the forced-choice pairs procedure, subjects choose one item each time whether they remember anything or not. Thus they have a 50% chance of choosing the correct item even when they are guessing. An estimate of memory accuracy based on this paradigm may be spuriously inflated by 50%.

Standing's [1973] study was the only one mentioned which attempted to take guessing into account in scoring for accuracy in the forced-choice pairs situation. Memory scores were determined by the formula: $S(Y - 2E)/T$, where S is the number of pictures shown, E the mean number of recognition errors, and T the number of recognition test pairs. This assumes that errors made by subjects occur at random positions in the test list, and that the same number of correct responses should be discounted as hits. Not satisfied completely with this guessing correction, Standing [1973], in another experiment, used a different procedure for correcting memory accuracy. In this experiment subjects were given continuous recognition lists of two, four, eight, 16 or 32 items. Subjects were allowed to choose only one old item in each list. This avoided response biases in subjects more or less likely to guess when not certain. In memory scoring, for each test list the number of errors occurring (Na) was divided by one less than the number (Na - 1) to give the estimated number of guessed responses (Ec). The number of items correctly remembered was calculated as S(T - Ec)/T. With two alternatives, memory was estimated at 97%, at four, eight and 16 alternatives at around 87% and with 32 alternatives at about 92%. These results were described as similar to results in the forced-choice pairs method. Standing also used a signal detection measure of discriminability (d′) as the recognition accuracy score. Though this method of scoring recognition is controversial, and will be discussed in detail later, it is far more sophisticated than the other corrections mentioned. Whereas correct responses and errors are lumped together in most correction formulas, in signal detection there is allowance for both correct responses to the old stimulus ("hits") and correct rejections of the new distractor. Also, a distinction is made between incorrect rejections of the old item ("misses") and incorrect acceptances of the new item ("false alarms"). Since subjects may be biased towards a particular kind of responding, corrections are made for this in signal detection scoring. In the forced-choice pairs situation, hits and correct rejections, or misses and false alarms, are made implicit by the single response, since to say "old" to one item of the pair is to say "new" to the other. But in any other recognition situation, it is important to know just what subjects mean by each recognition response. In forced-choice pairs testing, subjects are forced to choose an item as old even when both may appear either as old or as new items. There are many response biases which distort accuracy estimates. For example, Nickerson [1965] discussed the problem of bias in his subjects' tendency to respond "new" rather than "old" in his continuous list test procedure, and to make more accurate "old" than "new" responses. Shepard [1967] used the forced-choice pairs method of recognition to avoid Nickerson's difficulties with response biases. The forced-choice pairs method may force subjects to limit their responses so that there are no biases in the amount of responding given by subjects. The method has the larger problem, though, that there can be no estimate of how often subjects guess. Guessing corrections, based on probability estimates, will not give an accurate measure of individual guessing tendencies. For this reason, the forced-choice pairs method yields memory scores which indicate very little about what subjects remember.

Standing's [1973] finding that memory estimates in the forced-choice situation are similar to those in the continuous list situation led him to develop a mathematical prediction about recognition memory, such that the function relating items remembered to the number of items presented is a power law. The studies discussed in the next section reveal that the data about memory capacity of pictures are not consistent across tasks differing in similarity of test items to distractors, and that any empirical generalizations at this stage are unwarranted.

2.4.2 The Influence of the Distractor Set on Recognition of Pictures

The very high performance of subjects in the recognition studies described above cannot only be attributed to guessing strategies. The task may also be too easy for subjects, leading to ceiling effects in accuracy scores. A "ceiling effect" is a distortion of the distribution of scores on a task when the task is too easy for subjects. In this case, there are many perfect scores which skew the distribution of scores. A more difficult task would not produce the skewed distribution caused by the ceiling distortion, since there would be more differentiation in scores of subjects who perform well. The easy task assigns the same (perfect) score to many individuals who actually vary in their ability to perform the task.

It is surprising that so much recognition testing with pictures has ignored word recognition studies which show that recognition accuracy is a function of the similarity of the distractors to the test items [e.g., Bahrick, 1964; Underwood, 1949, p.512]. Using recognition distractors chosen on the basis of similarity to the test item in a study of memory for pictures, Bahrick and Boucher [1968] found that subjects tested immediately recognized 33% of the pictures, while two weeks later subjects recognized less than 23%. Bahrick and Bahrick's [1971] subjects recognized about 26% of pictures presented after two weeks. These findings show that the task was difficult, and subjects' memories were not completely accurate. These estimates of memory capacity are much lower than those found by researchers cited in the last section. In general, the studies which give high or low estimates of memory capacity vary in the kind of recognition distractors they use. The high-capacity studies reviewed above all used recognition distractors which were completely different scenes varying in every dimension from the originals. These distractors can be called "heterogeneous" because each is very different to every other. Studies which have yielded lower estimates of memory capacity have used "homogeneous" distractors, which are defined as being more similar to each other than the heterogeneous distractors.

Often homogeneous distractors are made by drawing different representations of the same objects, so that all items in the distractor set are from the same class. However, homogeneity is actually a continuum of stimulus similarity rather than a dichotomy between heterogeneous and homogeneous pictures. It appears to influence recognition by the simple rule: the more similar the items in a distractor set, the more difficult it is to recognize an old item amongst the new, other factors being equal.

The Bahricks have been most often credited with a move towards "homogeneous" pictorial stimuli. Bahrick, Clark and Bahrick [1967] tested subjects' recognition memory for 16 drawings of common objects (heterogeneous pictures) by asking them to choose test items from 11-forced-choice alternatives of similar pictures. The 10 distractors in each set were chosen from a set of 100 drawings of the same object (e.g., a cup), which were rated on a nine-point scale by 10 judges for their similarity to the test item. Two items failing most closely to each of five equally-spaced points on the continuum of similarity based on the ratings were chosen to make the pool of 10 distractors. These pairs represented first-, second-, third-, fourth- or fifth-degree similarity to the test item. In a recognition experiment in which subjects saw a test list only once, the distractors which were least similar to the test item were easiest to identify as "new". Using the same stimuli and distractors, Bahrick and Boucher [1968] presented the test stimuli for two seconds each without any special remembering instructions. One group of subjects was tested immediately and a second group was given the recognition task after two weeks. In the immediate group, recognition errors increased with similarity of distractors to their test prototype. This was not always true for the delayed groups. A later replication of part of this study [Bahrick and Bahrick, 1971] did find the predicted relationship between recognition difficulty and their measure of distractor similarity.

The Bahricks demonstrated that the similarity of the distractors to the test item does have relevance for estimates of the capacity of recognition memory for pictures. There are two problems, however, with their definition of similarity. In the first place, it was based only on ratings of similarity between prototypes and distractors. Account should also be taken of the similarity of distractor items to each other. For example, if distractors were more similar to each other than to the prototype, the prototype would stand out as more distinctive in the list. This is likely to influence subjects' response strategies in an unpredictable fashion. Certainly, there did seem to be occasional problems in the research mentioned, when the predicted correspondence between similarity and recognition did not occur.

There is a second and more serious problem. Pictures are complex arrays of visual detail, and can differ from each other in many ways, depending upon a momentary perceptual viewpoint. Certainly, judges can examine many pairs of pictures and respond to the relative similarity of pairs using a simple rating scale. However, we know little about the decision processes which the subjects use to produce the ratings. It may not be valid to ask subjects to compare stimuli differing in many ways, not all of which can be attended to at one time in order to make a useful rating. When so little is understood of subjects' judgments of similarity, it seems doubtful to use ratings which may be neither meaningful nor reliable. This has been a continuing problem in this area. Ratings procedures should be verified in a comparison with other measures of similarity based on physical properties of items.

Other studies using homogeneous distractors have given varying estimates of memory capacity. These have tried to control visual stimuli by using drawings and varying the similarity of distractors. Howe [1967] showed subjects sequences of 100 black and white illustrations of objects of a single category (e.g., ships) in a continuous recognition paradigm. Subjects recognized about 73% of items correctly. The false alarm rate reported was 18.6%, which is higher than that reported by Nickerson [1968], so it appears that Howe's subjects had more difficulty discriminating between test and distractor pictures. He also used a shorter presentation item than Nickerson or Shepard, 1.5 seconds per picture, which could have accounted for the lower recognition scores. Nevertheless, the capacity estimate in this study was lower than in the studies previously mentioned which used heterogeneous distractors.

Recently Tversky and Sherman [1975] have combined the forced-choice pairs technique with use of homogeneous distractors. Subjects were presented with 60 "dictionary-type" drawings and then shown 60 forced-choice pairs, where each item appeared with a similar drawing of the same item or items. Recognition was affected by both presentation time and interstimulus interval, ranging from a low of about 73% remembered at 0.25 seconds presentation time and 1.5 seconds interstimulus interval, to a maximum of about 93% with two seconds presentation time and three seconds interstimulus interval. Thus temporal variables as well as item similarity affect estimates of capacity for pictures in memory.

Another large-scale investigation used stimuli more homogeneous than the Bahrick pictures and found inter-picture differences in memory capacity. Goldstein and Chance [1970] used three types of black and white stimuli: faces, ink blots and snow crystals. Different groups of subjects viewed 14 stimuli of one type for two to three seconds each and then were tested for recognition either immediately or 48 hours later with a set of 84 pictures, of which 70 were new distractors. These pictures were from the same pool as the test pictures, although similarity was not controlled. The 84 pictures in the test set were presented sequentially and subjects had to make "old" or "new" responses to each item. About 72% of faces were correctly recognized, whether testing was immediate or delayed. 51% of ink blots were recognized immediately, but this dropped to 43% at 48 hours. Snow crystals were recognized very poorly, with only 37% remembered immediately and 30% after two days. Memory dropped significantly over the delay period for the latter two groups.

The study by Goldstein and Chance followed from an earlier study [Rock and Engelstein, 1959] which showed that subjects could have very high recognition accuracy in a homogeneous recognition task. This study was sophisticated in choice of distractor items. Rock and Engelstein asked judges to rate items in the pool of possible distractors for each picture on a 10-point scale of similarity to the original. The task was apparently too easy for subjects, as most had perfect scores, but this may have been due to the fact that only one item was tested at a time and distractors may not have been as well drawn as the originals. The choice of distractors on the basis of scaled similarity was a sophisticated fore-runner of the Bahricks' work.

Some homogeneous recognition studies have been greatly influenced by verbal studies in their choice of recognition distractors. This may be because one operational definition of homogeneity is that all pictures have the same class label. For example, one study has attempted to define similarity of recognition stimuli by controlling both the label similarity and visual similarity of recognition stimuli in pairs of photographs. Wyant, Banks, Berger and Wright [1972] asked pilot subjects to rate 650 pairs of pictures on the describability of the differences in the pairs and their visual similarity to each other. In testing, 219 pairs were chosen to represent three levels of verbal describability and of visual similarity. One item in each pair was presented as a stimulus. The other was presented as its distractor in a forced-choice pairs testing procedure. The authors found that there was a significant drop in percentage correct recognition with decreasing verbal describability of pairs, but not with their increasing visual similarity, when a 10 second rate of presentation was used for stimuli. However, there was a significant drop in recognition with increasing visual similarity with a three second rate of presentation. Wyant *et al.* interpreted this finding as evidence of some sort of visual encoding during the first three seconds of a picture's presentation, with verbal elaboration increasing in usefulness over the entire presentation.

Actually, it is difficult to imagine that the two types of ratings are mutually exclusive. Bahrick *et al.* [1967] did not specifically ask subjects to differentiate between verbal and visual similarity in giving their ratings. In the Wyant *et al.* study there is no check to determine whether pictures at one level of verbal or visual similarity differ on any other dimension. For example, the picture pairs which were rated as having differences very difficult to describe verbally were themselves classified with three levels of visual similarity. However, these pictures seem to be more similar visually, that is, more homogeneous, than the pictures classified as easy to describe verbally. Such ratings are difficult to interpret and difficult to use as an independent variable. It would have been better in this study to have measured similarity more directly. Latency to discriminate differences in the pairs of pictures could have been used as the visual similarity measure, while some measure of ease of verbalizing between pictures could have been used instead of the difficult ratings procedure. This might provide a more reliable and valid measure of the similarity of distractors.

Summary of Studies Using Adult Subjects

Studies of recognition of pictures using heterogeneous items as distractors have produced very high estimates of memory capacity. On the other hand, when distractor items are similar to the test stimuli, estimates of capacity are much lower. Memory capacity, then, is a function of the similarity of items to their distractors in recognition. No absolute measure of memory has been found.

It seems that recognition of pictures is not as good as the high capacity studies [Nickerson, 1965, 1968; Shepard, 1967; Standing, 1973; Standing *et al.*, 1970] have suggested. Homogeneous recognition testing methods have shown that recognition of pictures can be high or low depending on delays between presentation and testing and on the similarity of distractors used in testing. Some researchers [Bahrick and Bahrick, 1971; Bahrick and Boucher, 1968; Howe, 1967] have used heterogeneous lists with homogeneous items for the recognition test, while others [Goldstein and Chance, 1970] have used homogeneous items for all stimuli and distractors presented to a group of subjects. Table 1 summarizes the capacity studies reviewed in this section. The wide variation in percentage estimates of memory capacity suggests that no absolute measure of memory capacity has been found. Recognition memory for pictures appears to be a function of several variables in the testing situation. An obvious one is that recognition becomes more difficult with increasing similarity of items in the recognition test set. Until this source of difficulty is controlled, variables in the testing situation, such as list length and presentation time of items, cannot be studied to determine whether they affect memory accuracy, because the difficulty of the recognition task itself may interact with any experimental effects.

Developmental Studies

There are several developmental studies which do not clarify the capacity of memory question, but which have produced more sophisticated methodologies and a greater awareness of problems with the recognition paradigm than have most studies using adult subjects. Two studies compared estimates of both homogeneous and heterogeneous memory capacity in the same experiment.

The Bahrick recognition experiment has been used to test memory in five-year-olds [von Wright, 1973]. Subjects were given the 16 heterogeneous pictures such as those used in adult studies [Bahrick *et al.*, 1967] and asked to name them as they appeared. Subjects were given an oral recall test either immediately or two weeks after presentation, followed by a recognition test. In the recognition test, subjects were shown six items simultaneously and asked to point to an old item. Subjects were not required to make a response each time. The five distractors presented in each set represented five degrees of similarity to the stimulus. Sixteen sets contained an old item. An extra 16 sets were interspersed with the test sets, each containing homogeneous drawings of some new objects not shown previously, with a prototype and five distractors differing in five degrees of similarity from it. The 16 extra sets were used as a heterogeneous test, to see if subjects had any information about the pictures seen. Only two subjects made false alarms to any of the new distractor sets, and 99% of recognition

Table 1

Estimates of the capacity of memory for pictures in the recognition studies reviewed.

STUDY	STIMULI	METHOD	DELAY	MEAN (%)
		<u>Heterogeneous Distractors</u>		
Nickerson [1965]	200 black and white photographs	continuous list of 400 items	3 minutes	98.0
			16 minutes	87.0
Nickerson [1968]	100 items from Nickerson [1965]	same as above	1 day	78.0
			7 days	64.0
			28 days	40.0
			360 days	32.0
Shepard [1967]	612 coloured photographs	68 forced-choice pairs	2 hours	99.7
			3 days	92.0
			7 days	87.0
			120 days	57.7
Standing, Conezio and Haber [1970]	2,560 coloured photographs	280 forced-choice pairs	2 - 4 days	90.4
Standing [1973]	10,000 coloured photographs	160 forced-choice pairs	up to 5 days	99.7
Nelson, Metzler and Reed [1974]	60 pictures or photographs	60 forced-choice pairs	no delay	98.0
			7 days	85.0

Table 1 (cont'd)

STUDY	STIMULI	METHOD	DELAY		MEAN (%)
Homogeneous Distractors					
Howe [1967]	245 black and white drawings	5 continuous lists of 100 items	No delay		73.0
Bahrick and Boucher [1968]	16 drawings	11 forced-choice alternatives for each stimulis	No delay		33.0
			2 weeks		23.0
Bahrick and Bahrick [1971]	same as above	same as above	2 weeks		26.0
Goldstein and Chance [1970]	14 faces 14 ink blots 14 snowflakes	continuous list of 84 pictures for each stimulus category	Faces:	No delay	72.0
				2 days	72.0
			Ink Blots:	No delay	51.0
				2 days	43.0
			Snowflakes:	No delay	37.0
				2 days	30.0
Rock and Engelstein [1959]	1 shape	11 test alternatives	No delay		100.0
			3 weeks		100.0
			4 weeks		90.0
Tversky and Sherman [1975]	60 drawings	60 forced-choice pairs	No delay		83.0
Mandler and Day [in press]	12 or 36 shapes	12 or 36 forced-choice pairs	Meaningful shapes		90.0
			Random shapes		78.0

responses were to the correct sets in immediate testing. After two weeks over 83% were correct, and false alarms had increased. Homogeneous item recognition was only about 50% correct with immediate testing and 39% after two weeks. This is a controlled comparison of homogeneous and heterogeneous recognition, showing higher estimates of performance in the latter and giving a more complex view of recognition capacity. This estimate of homogeneous capacity is higher than that found by Bahrick and Boucher [1968], but subjects had the advantage of seeing each picture three times during presentation and naming the pictures as well. Also, subjects were not informed about the recognition task, though they were informed about the recall of picture names. The findings from this developmental study may have relevance for the problem of estimating the capacity of memory for pictures. Heterogeneous recognition testing yielded high estimates in comparison with results of a homogeneous test in the same experiment.

Another study compared homogeneous and heterogeneous information in two different recognition paradigms, the forced-choice pairs method and the continuous list method. Using coloured pictures of animals and people as stimuli, Brown and Campione [1972] used a continuous recognition paradigm with preschool children. Distractors were either homogeneous pictures, with the same characters posed in a different position, or new, heterogeneous pictures. Subjects judged each test picture as the same, similar or different to the original. Subjects recognized 95% of items correctly as being identical and 99% correctly as being similar to one seen previously. The false alarm rate for new items was less than 1%, so the children were not biased towards positive responding. In a second experiment, with delay before presenting the test list in a continuous fashion as before, correct recognition declined from about 92% at two hours' delay to about 78% after one week, a significant drop. Recognition for identical and similar pictures was the same. New items were consistently identified about 96% of the time. This is an extension of recognition methodology in an important direction. Heterogeneous recognition tests measure class recognition, while homogeneous recognition measures memory for a very particular instance of a class. Brown and Campione's recognition method, in which subjects judge whether a picture is similar to the original, probably approximates the natural recognition process more closely than other methods. More information is obtained from subjects, as they must identify general features of items in recognition and then determine whether the pictures are identical. The combination of homogeneous and heterogeneous information in this recognition task seems to be a useful approach.

Brown and Campione [1972] attempted to measure their subjects' response biases and guessing tendencies. Looking at accuracy rates as a function of the kinds of responses subjects made, they found that the probability of being correct when responding "identical" to the old stimulus remained high over delay intervals (0.94 to 0.97) but the probability of being correct when responding "similar" went from 0.92 to 0.75. Likewise, the probability of being correct when responding "new" dropped from 0.92 with two hours' delay to 0.75 with one week's delay. Accuracy decreased steadily over time for similar and new items, but subjects appeared to be using such a conservative criterion for accepting old pictures that accuracy of these judgments over delay was not affected.

In a third experiment, forced-choice pairs were used to test recognition. This allowed a comparison of continuous and forced-choice pairs testing methods in recognition. Half the pairs contained an original with a similar picture (the same animal or person in a different pose). An equal number of new pairs was interspersed in the test list also; in these, two poses of the same new character were seen. Again with this paradigm, there were very few false alarms; new pairs were described as new about 96% of the time. Of items not rejected as new, subjects had a probability of correctly identifying an old item of about 0.94 at two hours' delay and this dropped significantly to under 0.70 at one week's delay. The probability of correctly identifying an old item, given a response to the pair, was higher in this experiment than in the previous experiment where lists of single items were presented, although proportions of correct responses to old or completely new items were almost the same. Brown and Campione explained this by saying that the children presented with single items had a bias towards responding "similar" rather than "the same" when the character presented on testing looked familiar. This was presumably bypassed in the forced-choice pairs, where subjects were able to identify the old item on 85%

of all occasions. In both experiments, subjects never made "same" responses unless they were really certain of being correct; in both cases subjects were correct about 95% of the time they made an "old" response.

Due to the complexity of the various accuracy measures Brown and Campione [1972] discuss, it is difficult to evalute their comparison of the continuous list method and the method using forced-choice pairs. The usual response measures in these studies, proportions of correct and incorrect responses, reveal no differences in the two methods. Similarly, bias against positive recognition responses appeared to be the same in both studies. It is difficult, though, to generalize this finding to adult studies, where biases may be more subtle or varied. There was one difference found between the experiments: with forced-choice pairs, subjects were more likely to identify the old item, while in the single-item continuous list, subjects were less accurate about choosing old items once they had decided that the item was either the "same" or "similar". The tendency to respond "new" to a new item in a list of items declined in accuracy over time, but it was not reported whether this was true in the pairs condition. Because of the incomplete reporting of conditional probabilities in this experiment, caution must be taken in interpreting the results. But the study is interesting since it contrasts two recognition methods under extremely similar conditions. Numbers of distractors (both new and similar) and test items were the same for all groups, and initial presentation of the stimulus pictures was the same. However, the continuous list task was probably more difficult than the forced-choice pairs task, since subjects have to make an absolute judgment about each picture's familiarity rather than choose the most likely of two pictures. Brown and Campione suggested that in the continuous list situation, subjects favour a conservative strategy of saying the picture is similar to the original rather than identical to it. In the forced-choice pairs task, conservatism in criterion would presumably affect the subjects' tendency to choose a pair as familiar, but not the tendency to choose an old item as "identical", given that the subject has made a "similar" response to a pair. Unfortunately, means and standard deviations for "similar" responses were not reported for all conditions. If subjects employ different strategies in the two recognition methods, they may have differed in their tendencies to respond "similar" or "identical". There may be subtle differences in recognition responding when different paradigms are used to test memory.

Studies by Jean Mandler and her colleagues working with young children have led to improvements in quantifying similarity of distractors to the stimuli used in recognition. Using drawings of common objects as stimuli, Mandler and Stein [1974] created five transformations of each original picture to act as its distractors. Each distractor differed from the original in one of the following ways:

(1) the entire picture was reversed,

(2) three items were deleted from the array of 10 or so items,

(3) one item was made larger or smaller,

(4) the location of one or two items was changed, or

(5) one object was replaced by a "conceptually similar" object.

Mandler and Stein used two pictures, each of which was presented either as a scene or as an unorganized collection of items. Two sets of transformed distractors were made for each picture, transformations on organized or unorganized pictures. Subjects looked at the two pictures first in either the organized or unorganized version, describing them aloud, and then were shown two lists of pictures and asked to make a "yes" or "no" response to each one in recognition. For each picture seen previously, subjects were shown the item and its five transformations in a random order. Recognition was best when distractors were transformed by reversing items in the picture. Substitutions of items for similar items in the original led to distractors which were chosen as often as the originals themselves. Apparently subjects could not distinguish between them. Size and rearrangement transformations were recognized somewhat better. Probability of correct overall recognition was 0.70. Also, recognition of organized pictures was significantly better than recognition of unorganized pictures. This superiority was significant only when pictures were tested with size, rearrangement or deletion transformations. It appears as though substitution transformations were too difficult, while there was a near-ceiling effect with reversal transformations. These factors may have prevented the demonstration of consistent stimulus differences. Also, one of the pictures was better remembered than the other both when they were tested as organized pictures and as unorganized pictures. Thus the question of capacity of memory for pictures depended in this experiment on an interaction between the type of stimulus picture and the type of distractor variation. After the experiment, subjects were shown all possible pairs of pictures in a set and were asked to choose the picture most like the original as they remembered it. These similarity judgments were subjected to a multidimensional scaling technique (TORSCA). Recognition accuracy for each picture in comparison with every other in its set was also calculated as a d′ score based on a signal detection analysis and subjected to multidimensional scaling. The two analyses produced similar results. These analyses rely on Euclidean space assumptions about memory structure and hence have to be handled with caution, but they give further evidence that ease of recognition of a picture is related to the judged similarity of its distractors.

Stein and Mandler have shown that recognition is also relative to the physical similarity of items and their distractors which can be defined as the amount of change from the original in each distractor. In one study [Stein and Mandler, 1974], they investigated children's memory of orientation of geometric figures by presenting subjects with pictures containing coloured geometric shapes (e.g., a semicircle or a diagonal bar) and testing recognition memory with a list containing the test item and its distractors. In the experiment of interest here, the test picture contained three geometric objects. Distractors were made by transforming the original picture nine different times, each transformation containing one or more left-right reversals of the geometric shapes. Some distractors were quantitatively more different from the target stimulus than others. Subjects recognized 85% of the stimuli correctly. In general, distractors with one change from the original were more difficult to reject as new items than distractors with two changes, but no significant differences were found between distractors with two or three changes. Ceiling effects may have prevented the appearance of more marked differences.

In a replication of this study, Stein and Mandler [1975] used the same picture with its nine transformations as well as an analogous picture containing three realistic drawings similar in shape to the geometric shapes in the first picture. Nine transformations of this picture were made according to the same procedure as in the earlier study. Second-grade children had near-perfect performances on both pictures, while for kindergarten children, recognition improved as the number of changes in the distractors increased: most dissimilar transformations were easiest to recognize as new or different. This demonstrates again the importance of the similarity of items and their distractors on performance in the recognition task.

The studies by Stein and Mandler [1974, 1975] showed that similarity of a distractor to its test item can be quantified using a physical scale of amount of detail which is changed from the original to produce a distractor. This unambiguous definition of similarity has advantages over experiments in which subjects are asked to make ratings of similarity, since the similarity scale is reliable and can be used in a variety of experiments, despite differences in subjects' perceptions of the stimuli and the task. Ratings of similarity, on the other hand, do not provide an ultimate measure of similarity, since they rely on sampling characteristics of the raters chosen and on the particular demands of the rating task. In some experimental situations, the similarity of items as defined by raters may be more relevant to the task than in others, but since we know little about the rating process, these differences cannot be taken into account. The use of a scale of similarity based on physical changes made to a stimulus seems a potentially more meaningful instrument for manipulating similarity in order to study its effects on recognition.

Another developmental study has produced recognition distractors varying on four levels of similarity to the stimulus pictures, based on the amount of physical detail changed [Siegel, Babich and Kerasic, 1974]. Distractors were made by altering the prototype line drawings in quantitative steps. Each distractor in a forced-choice pairs paradigm was a variation from the original presented with it. Distractors could differ from the original in one, two, four or all details. Distractors differing in all details were actually heterogeneous pictures. Among the homogeneous stimuli and distractors, some were physically "more homogeneous" (i.e., more similar) than others. Siegel *et al.* found that fifth-grade subjects were able to reject more distractors which were heterogeneous or had four details changed than distractors differing in one or two details, though there were no significant differences between distractors with one or two details, or with four or more details changed. Latency to make a correct recognition response was greater for items differing in one than two details, and for those differing in two rather than four details, showing that it was more difficult to recognize items in similar pairs. Recognition was still very high in this study; 73% of old items were correctly recognized when distractors differed in only one detail, while 92% of old items were recognized in pairs with the most different distractors. This confirms findings of previous studies, showing high capacity in heterogeneous pairs and lower capacity in more homogeneous pairs.

Siegel *et al.* changed different types of details in their distractor items, while Stein and Mandler [1974, 1975] used only left-right item reversals in making changes. Since Mandler and Stein [1974] found that some types of changes lead to more detectable differences than others, it may be wise to use only one type of change in creating a scale of physical similarity. However, this creates the problem that only information of a particular type is being tested, not memory for the whole stimulus. Both these approaches must be used, so that memory for particular information as well as for a combination of different kinds of information can be understood in relation to each other.

2.4.3 Conclusion

Studies of capacity of memory for pictures using heterogeneous recognition sets seem to overestimate the amount of pictorial information which can be retained. When a picture is tested for recognition with distractors which are homogeneous, recognition accuracy is a direct function of its similarity to its distractor items. Recognition of a picture is poor when the pictures from which it must be chosen are similar to each other. Heterogeneous recognition studies give capacity estimates of 90% or more in immediate testing of memory for pictures, while homogeneous recognition studies give estimates as low as 33%. Obviously, an estimate of capacity is relative to each experimental situation in which memory is tested.

The reason for a difference between estimates of memory capacity in heterogeneous and homogeneous recognition tasks is probably due to a methodological peculiarity in the recognition paradigm. McNulty [1966] noted that subjects can perform adequately in a recognition task with only "partial learning". This can easily happen in a picture recognition task, when subjects remember only a small detail of a picture, which is enough information to allow the subject to reject all the distractors. In this case the subject is credited with remembering the picture when in fact only a minor detail about the picture is remembered. Subjects in heterogeneous recognition tasks may rely on partial information for every response, while this is much less likely in homogeneous recognition tasks, where much detailed information must be used in order to discriminate between the test items. Partial memory may well be a normal part of remembering, but in studying recognition it must be strictly controlled. Homogeneous recognition studies provide the only paradigm in which spurious success due to partial memory can be minimized, so that an informative study of the capacity of memory for pictures can be undertaken.

Not all the studies using homogeneous recognition stimuli have minimized effects of partial memory in their recognition scores. Distractors which are totally different drawings to the originals [e.g., Bahrick and Boucher, 1968] may vary in many details from the originals, since their similarity is determined by using judges' global, subjective ratings. Thus, theoretically, a subject could remember only a small detail of the original, and eliminate every incorrect distractor on the basis of this, despite the subject's poor general memory of the picture. In some of the developmental studies mentioned [e.g., Siegel *et al.*, 1974; Stein and Mandler, 1974, 1975] the original picture is changed slightly to produce the distractors. This provides less opportunity for partial memory to influence recognition responses, inflating estimates of memory capacity.

However, the partial memory problem cannot be eliminated completely. It is intrinsic to the recognition paradigm. Thus recognition scores will never yield an absolute measure of memory. Rather, recognition gives a measure of performance on a memory task relative to the difficulty of the distractor sets used in testing. A true measure of storage capacity must await a full understanding of the partial memory phenomenon in the performance of subjects in a recognition task.

2.5 THE CHOICE OF A METHOD FOR MEASURING MEMORY FOR PICTURES

2.5.1 Recognition Versus Recall

Recognition testing does present methodological problems in its susceptibility to the effects of partial memory and guessing strategies, which lead to variable and often inflated accuracy scores. Recall also suffers from a serious disadvantage in its reliance on the appropriateness of subjects' retrieval strategies. For this reason, the material retained may be recalled well or poorly depending on the cues the subject is given in the test situation. In recognition the cues are, to a greater extent, under the experimenter's control (e.g., the similarity between the old stimulus and its distractors). In free recall, the cues are more subtle and little is known about the process which aids the subject in recalling the first stimulus and how this contingency affects the recall of the next items. Because retrieval is more idiosyncratic in recall than recognition, there is a greater chance that there will be information which is not accessible with a free recall procedure but which may be accessed in recognition. When so little is understood about processing in memory for pictures, the method which yields the most information about memory should be used to study its capacity and the variables which influence it. Thus recognition is the method of choice in this study. The investigation of memory using a recall paradigm must await the development of more suitable methods than the techniques of drawing pictures or writing down lists of words, which were criticized earlier.

As mentioned earlier, there is some doubt that memory performances in recall and recognition tests are operational definitions of the same memory process. Data from verbal studies [see Kintsch, 1970] suggest that free recall leads to a smaller number of correct responses than recognition in tests of memory for word lists, and Kintsch [1970] has used this as evidence for hypothesizing that recognition and recall are distinctly different processes, recall being more difficult because of a greater complexity of retrieval operations involved. Although most studies which have measured memory for pictures under recall conditions, where subjects have to draw pictures they have seen from memory, have not yielded measures of the capacity of memory, accuracy in these studies seems to be low, much lower than in the recognition tasks discussed. Also, verbal recall of pictures is inferior to picture recognition both when recall is by writing down one-word labels for pictures seen [Bahrick and Boucher, 1968] and when recall is by brief verbal descriptions of photographs [Standing, 1973]. However, there are methodological difficulties in comparing two different measuring instruments, and recall and recognition scores are not comparable to each other in many ways. In recognition, there are controversies over corrections for guessing and use of signal detection methods. In recall, the total number of items recalled is usually scored, but sometimes clustering scores are used. There is little information about whether different techniques and scoring systems measure the same aspects of memory. It may not be valid to compare scores obtained under one method with those obtained under another, since the memory scales used in each test are not comparable.

Other researchers have not tried to compare recognition and recall directly but have tested whether the same subjects do well or poorly in both paradigms. Bahrick and Boucher [1968] found low correlations between recognition of pictures and their verbal recall in different experimental groups. They concluded that recognition and recall techniques differ in their use of retrieval processes in memory. However, Bahrick and Bahrick [1971] found low correlations between recognition performance on a test of noun labels denoting pictures shown and a pictorial recognition task presented with the same items after the verbal task. This suggests that verbal measures and pictorial measures differ. Thus the relevant comparison between recognition and recall in pictures should not include any verbal responses.

Some studies have shown that the relative efficiency of recall and recognition in memory performance with pictorial stimuli is a function of situational demands. Frost [1972] found that subjects expecting a recognition task performed better than subjects expecting a recall task, but subjects recalled the same amount of pictures no matter what instructions they had been given. All subjects were given the verbal recall task followed by a forced-choice pairs recognition task, and in general performed better on the recognition task. Tversky [1973, 1974] found evidence for a more general instructional effect; subjects perform best on tests when they have received appropriate instructions. Subjects who expect a verbal recall task recall more pictures than subjects who expect a recognition task, and *vice versa*. In both experiments, recognition was generally found to be higher than recall. Tversky's [1973] study was an improvement on Frost's, because subjects were given the recall and recognition tests in a counterbalanced order. Only recall was affected by order of testing.

There is no clear evidence to suggest that recall and recognition are different processes. However, it is known that recognition scores are usually higher than recall scores, and that expectations about recognition or recall lead to different memory strategies and performance.

Theoretically, it may be possible to vary recognition distractors so that recognition is as difficult as a recall task when pictorial stimuli are used. This would not be achieved by equating the cues provided in the two tasks, because subjects in the recognition task are by definition given cues which are not provided in free recall. Rather, the difficulty of the recognition task can be manipulated by creating a set of distractors sufficiently similar to the target picture. The poor performances in the recognition and recall tasks would arise from two different sources, however. In recognition, difficulty is increased as more pictures which can be confused with each other are shown, so that interference may be the main problem in retrieval. In recall, difficulty is increased by removing contextual cues so that subjects must attempt to retrieve information from memory without many external aids. Thus an experiment which successfully equates recognition and recall in task difficulty does not ensure that memory processing is the same in both paradigms. Until more is known about memory processes involved in both recognition and recall paradigms, it will not be possible to determine whether these methods measure the same memory processes. In the study of memory for pictures, recognition is the most appropriate method because it has the potential to measure nonverbal information without the requirement that subjects use any complex motor skill.

The problems of partial recognition and guessing strategies have not been adequately dealt with in studies to date. A new recognition paradigm will now be described which attempts to control these factors to produce a useful measure of memory accuracy.

2.5.2 The Exclusion Set Method

Rationale

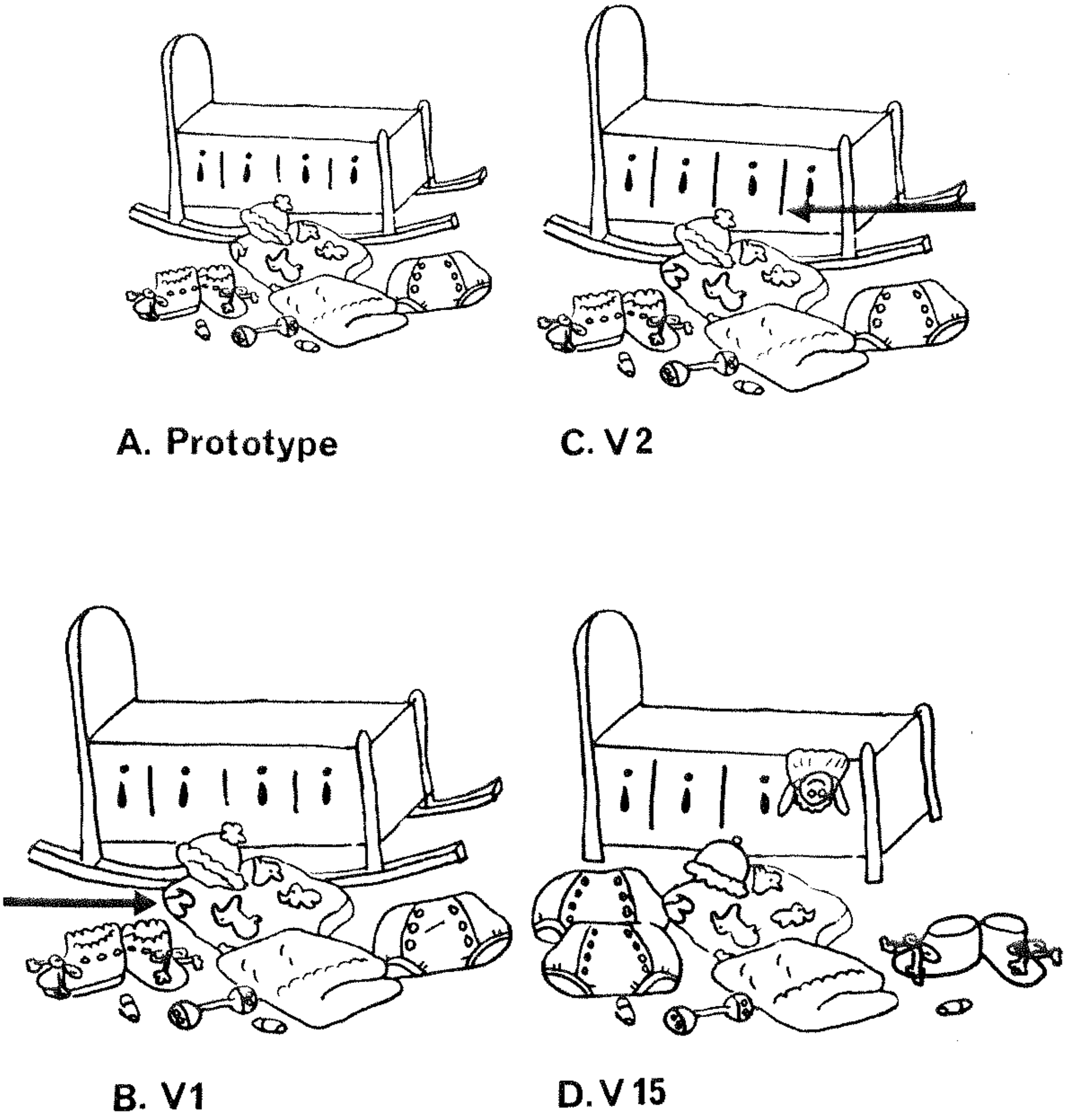
The exclusion set method is based on the idea that the only judgment that a subject with a partial memory can make about a recognition item and be confident of his accuracy is that it is *not* the original. The judgment that it *is* the original must always be a guess. For example, a subject with only a partial memory for a picture is usually forced to choose between the stimulus picture and a set of distractors in a recognition task. The subject should be able to exclude any distractor whose structural description in perception [cf. Sutherland, 1968] is incompatible with his incomplete description of the stimulus in memory. If the subject were forced to select a single item as "old", he would have to guess between those items which do not conflict with what he remembers. This is presumably the subject's usual procedure in a recognition experiment, but it discards much information about the subject's memory. Instead, the exclusion set method does not force the selection of a single item as "old". The subject is asked only to exclude those items which he judges to be different from the original. The size of this set provides a measure of the detail about the original which the subject has retained in memory.

Implementation of the Method

The success of the method depends upon the choice of the distractor set, which must be sufficiently homogeneous to detect the amount of detail lost from memory. Conventionally, "homogeneous" distractor items [e.g., Bahrick *et al.*, 1967] are similar to the original globally, but differ from it in every concrete detail. The fact that recognition performance is variable and sometimes high in experiments using this material (see Table 1) suggests that the material used is not sufficiently homogeneous. Instead, in the present study, distractor items were used which were *identical* to the original in all but certain selected details. By appropriate selection of these changes, it was possible to guarantee that confusions will occur and hence to provide a highly sensitive measuring instrument.

In order to create a recognition set which could discriminate between subjects with varying levels of memory for a stimulus, a set of distractors was created which varied in their similarity to the original stimulus. This constituted a sequence of items progressively more different from the picture which was to be tested. Line drawings representing complex scenes with separate, identifiable parts were used as the stimulus prototypes. A distractor set was made by making successive, cumulative changes to each prototype. The first variation in the set, most similar to the prototype, was changed in only one detail from the original. The second variation created retained this change, and a second detail was altered so that the picture differed in two details from the original. The last variation had been subjected to many cumulative changes, and was most different to the prototype. An example of a distractor set containing an original prototype and variations changed in one, two or 15 details is shown in Figure 2. The distractors can be seen as ordinal points on a scale of increasing change in physical detail from the prototype.

Any picture in the set which is excluded during recognition testing contains a finite number of details which are incompatible with an accurate memory of the stimulus. The distractor sequence can be thought of as sampling the picture until a detail is changed which is incompatible with the subject's memory. The poorer the subject's memory of the picture, the lower will be the probability that a given changed detail will be detected and lead to a rejection of a distractor. In this case, a large number of details would have to be sampled before an incorrect picture could be rejected, and the subject with a poor memory would be given a small exclusion set score. If a subject's memory were accurate, there would be a high probability that a detail sampled by a distractor would conflict with the memory for the prototype, and the exclusion set score would be large. Because of the way the test



A. Prototype

C. V2

B. V1

D. V15

Figure 2. An example of a prototype picture (Picture 2, Experiment 1) and its first (V1), second (V2) and fifteenth (V15) distractor variations. Arrows point to the first two details changed in the set. (Each picture shown in the test sets was the same size.)

sets are created, the distractor variation most similar to the prototype which the subject can exclude gives a good estimate of the amount of detail about the prototype which the subject has remembered. This rejected distractor and all distractors above it in the test set, less similar to the prototype, form the basis of the subject's "exclusion set" which is the memory score derived from the method. (This assumes that subjects who reject a distractor containing erroneous details would be able to reject all other distractors containing the same details; that is, all distractors higher on the test set.) All pictures which the subject cannot reject on the basis of a changed detail are then said to be in the subject's "inclusion set". The better a subject's memory, the larger will be the subject's exclusion set and the smaller the corresponding inclusion set. Recognition testing in the paradigm proceeds to find the cutoff point in the test set for each subject at which the subject can reject all higher variations and cautiously accept all lower variations from the prototype.

The problem of partial recognition is controlled in this paradigm. The details chosen to be varied in the distractor set are assumed to be a random sample of possible changes to the prototype. The partial recognition effect is produced by the subjects' ability to exclude a picture on the basis of a small amount of remembered detail. There are two kinds of details which can be changed in the distractor sets, salient details and non-salient details. If there are salient details which all subjects remember, then partial recognition will have a huge effect on memory scores in the exclusion set method if these salient details happen to be changed in the distractor set. For example, if the first variation in the test set is changed by altering a salient detail, all subjects will be able to exclude all the distractors. There will be a ceiling effect in exclusion set scores. For this reason, salient details are avoided in distractors for the exclusion set method. Partial recognition may also be due to each subject's memory for idiosyncratic detail. Thus if details from a picture are randomly sampled, occasionally a subject may be able to use partial recognition to exclude a distractor very similar to the prototype even though his memory is poor. But this should be a rare occurrence and, with a large pool of subjects, the effect on average memory scores will be minimal. The random sampling of non-salient details from the pictures to create distractors ensures that partial recognition does not appreciably affect exclusion set scores.

A potential problem with this method is that subjects can reject distractors on the basis of wrong information. The validity of the exclusion set method relies on the assumption that subjects exclude pictures because of specific details which clash with their memories. The problem is overcome by asking subjects to make confidence ratings about their rejections and to state reasons (i.e., name details) for their rejection of a distractor. Subjects who give ratings of low confidence or who give incorrect reasons for rejecting a distractor are not scored as having correctly rejected that distractor. This procedure also ensures that subjects who are guessing are not given credit for rejecting a distractor.

The actual procedure involved in testing recognition with the exclusion set method is discussed in the next chapter. From this outline of its assumptions and basic methodology, it does appear to be as difficult as a recall procedure but does not have the same scoring problems, such as subjects' production difficulties in drawing. Problems which have been common with the recognition paradigm, partial recognition and lack of control over subjects' guessing strategies, have been subjected to experimental control. In the next two chapters, this new recognition methodology is used to examine effects of delay and instructions to remember on subjects' memory for pictures.

CHAPTER THREE

LONG-TERM CHANGES IN MEMORY FOR PICTURES

3.1 INTRODUCTION

There are as many estimates of the capacity of recognition memory for pictures as there are studies, as Table 1 amply demonstrates. One uncontrolled factor which may be contributing most to the diversity of these estimates is that of partial memory, discussed in the last chapter. Since in most recognition studies, subjects can choose the correct item when little is remembered about the stimulus, and since the extent to which this will lead to inflated memory scores will largely depend on the similarity between items in the recognition test, no studies properly estimate the adequacy of subjects' memories for a single picture. The partial memory effect must be minimized before any estimate of capacity for a list of pictures can be calculated.

Many of the studies measuring the capacity of memory for pictures, which were reviewed in the last chapter, have shown that memory decays over time. The time intervals studied in experiments using heterogeneous test items have usually been longer than those employing homogeneous test items, and have found that memory remains accurate for long periods of time, with a slow decline. However, it was pointed out earlier that this result may be due to ceiling effects in the data from heterogeneous studies. This may lead to an underestimation of the effects of delay on memory for pictures. One heterogeneous study [Nickerson, 1965] found that items tested later in the memory list were more poorly remembered than those tested earlier, but this may have been due to interference rather than to forgetting over time, since there were more intervening items before the pictures tested later. None of the studies reviewed showed memory loss after one day, but recognition of pictures has been shown to fall significantly after two days using homogeneous test items (Goldstein and Chance, 1970) or seven days using heterogeneous test items [Nelson *et al.*, 1974]. Various studies have shown a decline in memory over various times up to a year, but have not indicated which delay groups differ significantly from each other [Bahrick and Boucher, 1968; Nickerson, 1968; Shepard, 1967; Standing *et al.*, 1970]. A methodological problem in some studies has been the use of the same group of subjects tested at different times [e.g., Nickerson, 1968]. In this situation, subjects may be getting more practiced with the task itself on their second test and perform better than they might have, had they been tested for the first time after a delay period. Because of the differences in various studies in use of delay groups, reporting of statistics, and similarity of items in the test set, there has been no conclusive study of the course of memory decline in recognition of pictures.

To attempt to rectify this, the experiment described in this chapter examined the decline in recognition memory for several different pictures over two months, in a paradigm which minimizes inflated accuracy scores due to partial memory and guessing. Different groups of subjects were tested at each delay interval. The effects of partial memory were controlled in the new testing method, the exclusion set method, by defining the details which can be remembered about a picture and used in a decision to reject an unfamiliar picture in recognition. The method is sensitive to fine differences in subjects' memories even when subjects are very accurate, and minimizes the influence of partial memory strategies on correct recognition choices when scores are averaged over a group of subjects. Experiment 1 was carried out to evaluate the hypothesis that memory capacity is very high for pictures by examining the amount of detail about an individual picture which subjects can remember. By testing recognition after varying periods of delay, the changes in memory capacity over time were examined to see if capacity remains high or whether remembered detail is quickly forgotten.

3.2 EXPERIMENT 1: EFFECTS OF DELAY ON RECOGNITION

3.2.1 Method

Choice of Delay Intervals: Pilot Testing

In order to select the delay intervals to be used in the experiment, a pilot experiment was carried out. From the pool of first-year students available, 25 volunteered for the experiment. Five subjects were assigned randomly to each experimental group, which consisted of an immediate-test group and groups given the recognition task after four hours, eight hours, one day or two days. These time periods were chosen on the basis of previous studies and covering what seemed a reasonable time span. Basically the same procedure was employed in the pilot experiment as in the experiment proper described below, except that the five stimuli used were different items drawn from the same stimulus pool. Subjects were given a single practice trial and then told to look at the test pictures which would be shown for 10 seconds each and to try to remember them, as their memory would be tested later.

There were only small differences in exclusion set scores in the five groups tested. There was, however, an indication of a reminiscence effect: memory declined steadily over the first day, but for every picture there was a slight increase in memory from one day to two days' delay. It was decided that memory probably decays little in a single day, but that there may be an increase in memory from the first day to the second day after stimuli have been shown. For the main experiment, a no-delay group was compared with groups delayed in memory testing by one, two, seven or 60 days. The choice of a group given one week's delay replicates a condition used in previous memory experiments, while the two months delay condition provided the longest practicable interval in order to measure long-term forgetting.

Design

The exclusion set method was used to test memory decline over delay periods of one, two, seven and 60 days. Different groups of subjects were tested for each delay period, so that subjects' experiences with the task itself did not affect their memory scores.

Subjects

First-year Psychology students participated in the experiment as part of a course requirement. Seven males and seven females were assigned randomly to each of the five delay groups, so that there were 70 subjects in all.

Stimuli

Six complex black and white line drawings were chosen as stimuli from a set of 20 pictures adapted from children's books. Each picture contained approximately 10 items in an integrated scene. Pictures chosen were those which were neither too easy nor too difficult for subjects, having displayed little variance in accuracy scores in a second pilot experiment. In this pilot experiment, 25 subjects had been shown five of the 20 pictures in a random order and had been tested immediately with the recognition task. Because subjects could not remember more than five pictures at a time (judging from preliminary testing), the 20 pictures were divided into four sets of five pictures, and each of these sets was presented to five subjects. From the recognition scores, four pictures remembered adequately by all five subjects in a group were chosen for use in the main experiment. A further two pictures were chosen as practice stimuli. Memory performance was slightly more variable with these pictures. Of the 20 pictures, 12 were omitted solely because at least one subject had a perfect score or null score in the recognition test.

Distractor sets were constructed for each stimulus. Each test stimulus was used as a prototype, and 15 variations of each prototype were used as the distractors. Variations were made by successive and cumulative alterations of the prototype. The prototype was drawn on transparent plastic with a felt pen and photographed. It was then altered by deleting or adding lines, changing the shape of an item, or moving a whole item to a different position and rephotographed. Changes considered salient to the whole structure of the pictures, as determined by features most often remembered by subjects in the preliminary study of picture reproduction (Chapter Two), were avoided. The 15 successive changes were made and the results photographed to make a test set comprising 16 items. This could be described as an ordinal scale of increasing change from the prototype, so that variation 15 was least like the original. An example of one of these sets (for Picture 2) was shown in Figure 2. The prototypes and their variations were reproduced as 35 mm slides. There were four test items, one practice item and a "baseline" item given after the practice item in order to gauge the normal ability of each subject for the purpose of equating groups. The other three test pictures, the baseline picture and the practice picture are presented in Appendix 1.

An attempt was made to scale the variations in each set using ratings of similarity. Each picture in the test set was presented with its neighbour above it in the test set (except, of course, the last variation) and judges were asked to rate the similarity of the items in the pair on a seven-point scale from "most similar" (1) to "least similar" (7). Raters were seated before a rear projection screen, onto which the pairs of pictures were projected, and given two photographic prints which were identical copies of the prototype being tested. They were told that the two pictures were identical and hence should be given a rating of "1". This anchored all raters' responses of similarity to one end of the scale. It was not feasible to show raters all pairs of pictures in each test set because of the numbers of pairs involved, so pairs of adjacent variations in each set were shown and the single differing detail in each pair was pointed out to raters by the experimenter. Only five raters were tested on this task before it was abandoned. They all complained that it was difficult to make ratings about pairs which differed on so many different dimensions. A typical comment was that pairs of items were similar in some ways and different in others, and that a single criterion was unsatisfactory. Thus the ratings of similarity seemed to be very difficult for the judges, and it was not clear that the ratings would have had any validity as a unidimensional scale of similarity. Because of this, items in the test sets cannot be described as points on a linear scale of similarity to the prototype. Since points on the scale are not absolute indices of amount of change from the prototype, the corresponding levels of variation of pictures in each set may not be comparable. Thus the *i*th variation of one stimulus prototype may not be as similar to it as the *i*th variation of another prototype.

Apparatus

Two Kodak Carousel slide projectors, Model S-RA, were used to project pairs of pictures side by side onto a rear projection screen in front of the subject. The projector on the subject's right was used to present the stimulus prototypes. The timing of the presentation of each stimulus slide and the interstimulus interval were controlled by an electronic timer which operated a shutter placed in front of the lens on this projector. Pairs of test pictures were presented by means of two projectors. Both slides of a pair were presented simultaneously. The time of presentation of each pair was paced by the subject's recognition decision. Each slide projected onto an area 26 cm by 17 cm. The images from the two projectors were 10 cm apart. Each picture subtended a visual angle of 9° by 13° with subjects seated 75 cm away from the centre of the screen. The experimenter sat beside the subject at a table with the controls for the projectors. Polaroid filters were placed in front of each lens to reduce brightness. The room was dimly illuminated.

Procedure

The test procedure for each picture was as follows: Subjects were seated before the screen and told by the experimenter,

> "The purpose of this experiment is to help me get some idea of how our memory for pictures works. I'm going to show you a set of pictures. Then later I'll show you a whole series of very similar pictures to see how well you can recognize the original pictures among a lot of similar ones."

The practice item was shown on the right side of the screen for 10 seconds. After its presentation, the subject was told,

> "Now I'll show you some pictures very similar to it. For each one I want you to decide whether the picture is actually the original one. Use this card which demonstrates

seven possible ratings you can give to each picture."

The subject was handed a card giving seven rating responses:

(1) I am positive it is the original.

(2) I am fairly sure it is the original.

(3) It might be the original.

(4) Not sure.

(5) It might not be the original.

(6) I am fairly sure it is not the original.

(7) I am positive it is not the original.

(Pilot testing had shown that five-point ratings were not adequate in discriminating between subjects' responses, so the seven-point scale was used.) Subjects were also told,

> "Don't guess when you are not sure. Ignore specks and colour variations in the slides, and don't worry if pictures look pretty much the same. I am just picking them at random with this machine."

This last instruction was added to overcome the problem of subjects' strategies in guessing how the stimulus sets were put together and to warn them of the homogeneity of the test items.

Recognition testing was carried out in two phases. In Phase One, subjects were presented with the items in a particular test set one at a time in a random sequence, and subjects had to give confidence ratings about each item. To determine whether subjects had failed to reject any distractors merely because they had failed to attend to a critical detail in a picture, Phase Two testing was carried out. In this procedure, pairs of pictures were shown which were adjacent variations in the test set differing in only one detail. The experimenter asked subjects if they could reject either of the pictures in each pair on the basis of the difference. One detail in each pair was correct, i.e., it occurred in the prototype, and one was incorrect. The procedure in both phases is described more fully below.

Phase One: Subjects were presented with the 16 pictures in the test set in an individually randomized order. They were asked to give seven-point confidence ratings for each picture. They were told to make ratings of "4" when they were uncertain about any picture. When subjects rejected a picture with a rating of "7", they were asked to give a reason for the rejection. When subjects rejected a distractor with a "7" rating and gave a correct reason for the rejection, they were not shown any variations higher in the set which contained the same detail subjects had just rejected. This method shortened the set of test pictures for each subject, and made testing less tedious. Thus, all subjects did not see the same number of pictures in Phase One testing. If subjects were not able to reject any distractors with a "7" rating, or could not give a correct reason for so doing, they were shown all 16 test pictures. However, a correct "7" rejection given by a subject to any of the pictures in the test set immediately led the experimenter to reduce the size of the test set shown to that subject, in the manner stated. Every subject saw the prototype and all variations closest to it which the subject could *not* reject for the correct reason. To a certain extent, this meant that subjects with accurate memories saw fewer pictures than those with poor memories. However, this did not always occur since the order of the pictures shown in the set was randomly determined. Thus, even if a subject had a perfect memory for a picture, any number of items randomly chosen from the test set might be shown before the subject would be presented with the lowest variation which he could reject correctly.

Phase Two: After the initial presentation of the test set, subjects were shown pairs of pictures in the set to see if they could reject any lower distractor variations when given a second opportunity to notice the critical details. One of the remaining items in the test set (which subjects had not rejected correctly in Phase One testing) was randomly chosen and presented to the subject along with the distractor variation above or below it in the stimulus set (determined at random). The single difference between these two pictures was pointed out to subjects, and they were asked if they could reject either one on the basis of this difference. Instructions to subjects were:

> "Now I'm going to show you some of the same pictures again, two at a time. They will be different in some way, which I'll point out to you, and I want to see if this helps you to reject a picture on the basis of this difference. It might jolt your memory. You might not reject either one or

> you might reject both occasionally. Don't guess if you're not sure. It doesn't matter what you've said before."

For each pair, subjects could: (1) make no response, or (2) reject one of the pictures in the pair and give it a rating of "5", "6" or "7". (On four occasions a subject rejected both pictures. No response was scored.) Once again, if a picture was correctly rejected with a rating of "7", no pairs containing higher variations were presented.

After the practice trial all subjects were presented with the baseline test item. The procedure for this item was the same as for the practice item. Afterwards, the four test stimuli were shown successively in an individually randomized order, with 10 seconds presentation time and approximately 10 seconds interstimulus interval. No-delay subjects were told that the four pictures they were to remember would be shown one after the other and tested immediately afterwards. Delayed-testing groups were told that they would be shown the pictures and were told that they were to remember them "for next time". Subjects were aware of the delay period involved before their second testing. No-delay subjects were given the entire procedure in one sitting, taking approximately 60 minutes. Delayed-test subjects were tested with the practice and initial pictures and then presented with the test stimuli, a procedure taking only 20 minutes. Upon returning for the second session of the experiment after one, two, seven or approximately 60 days, the initial instructions were repeated to subjects and then the testing procedure was carried out as for the no-delay group. This took about 45 minutes. There was an attempt to assign subjects randomly to delay conditions, but some subjects were only available at certain times and were assigned to convenient groups. The delay intervals were kept as precise as possible, but subjects in the 60-day group were asked to return as close as possible to two months from the original date, and all returned after 56 to 60 days.

After recognition testing, all subjects answered the following questions:

(1) Did you try to use any strategy for remembering the pictures? Did you think about the pictures over the past day/two days/week/months?

(2) Was the original picture in each set you saw today?

(3) Did you think any pictures in each set were repetitions?

(4) Could you figure out which picture was the original one in each set even if you couldn't remember? What did you think if an item came up time and time again in a series or hardly ever appeared?

3.2.2 Results

Scoring

The measure of recognition accuracy used in the present experiment was that of exclusion set size. This was conceived as the number of items given a rating of "7" with a correct reason. However, since some of these items were not presented directly to subjects, the score was actually 16 minus the number of the lowest distractor variation given a correct "7" rejection.

Data and Results

Means and standard deviations of exclusion set scores for each delay group for each of the four pictures are shown in Table 2. Mean scores are also plotted as a function relating accuracy of memory to delay in Figure 3 for all pictures combined, and in Figure 4 for each picture separately. In general, memory accuracy declined over time. Accuracy scores for two of the pictures were related to delay by a decreasing function, while the other two pictures seemed to show a reminiscence effect, since there was a slight increase in memory from one to two days' delay.

To test whether the different delay groups differed in initial memory ability, a one-way analysis of variance was carried out on the exclusion set scores for the baseline item. Means for the groups were: no delay, 7.857; one day, 7.214; two days, 8.071; seven days, 7.643; and 60 days, 7.714. There was no significant effect ($F < 1$, $df = 4,60$, $p > .05$), indicating that groups did not differ in initial ability. Because of this, it was not considered necessary to carry out an analysis of covariance on test scores.

Table 2

Means and standard deviations of exclusion set scores for each delay condition and each picture in Experiment 1.

Picture	Delay Group					Total
	No Delay	1 Day	2 Days	7 Days	60 Days	
1. $\bar{X}$	6.3571	4.9286	6.6429	4.5000	2.0000	4.8857
S.D.	2.9770	4.1964	2.9249	4.3279	2.4179	3.7438
2. $\bar{X}$	8.7857	7.1429	5.9286	4.3571	2.4286	5.7286
S.D.	2.5774	3.6132	3.6472	3.8151	2.5333	3.8820
3. $\bar{X}$	7.9286	6.4286	6.5714	5.7143	2.2143	5.7714
S.D.	3.4073	3.6314	3.0813	2.6726	3.3553	3.6956
4. $\bar{X}$	8.4286	5.4286	7.2143	6.3571	1.7857	5.8429
S.D.	2.6808	3.6314	2.5474	4.8136	2.1901	3.9366
Total $\bar{X}$	7.8750	5.9821	6.5893	5.2321	2.1071	
S.D.	2.9973	3.7731	3.0258	3.9680	2.5984	

The baseline scores were also subjected to an analysis of sex differences. Scores for all delay groups were pooled so that a t-test comparing independent means scores on the initial item for all female subjects and all male subjects could be carried out. No significant difference was found ($t = 0.084$, $df = 68$, $p > .05$ for a two-tailed test).

A two-way analysis of variance was carried out treating the repeated measures on the picture factor as a random effect.[1] The analysis of variance summary table is shown in Table 3. There was a significant effect of Delay, but not for Pictures, or for the interaction of Pictures with Subjects within Groups. Thus it may be concluded that there was a decline in memory for pictures over time, but there were no differences in memory for any of the four pictures at any delay. Scheffé comparisons were carried out after the analysis to test which

---

[1] Since the analysis of variance is considered to be robust with respect to its assumptions of normality and homogeneity of variance of the samples [Box, 1953], and since appropriate nonparametric tests which do not rest on these assumptions (which are difficult to test) were not available to provide any alternative analysis, parametric tests are used throughout these studies because of their greater power.

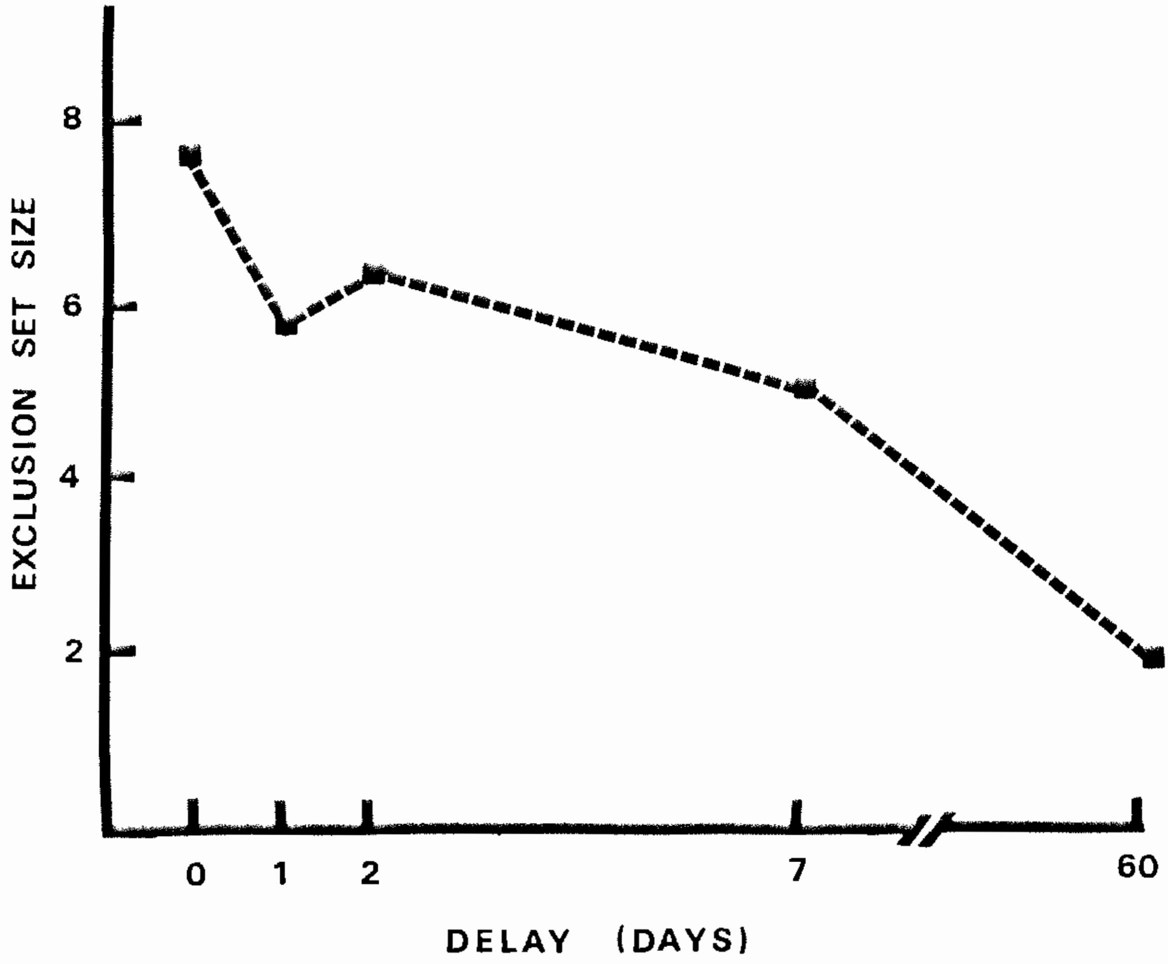


Figure 3. Recognition accuracy (exclusion set size) as a function of delay in Experiment 1.

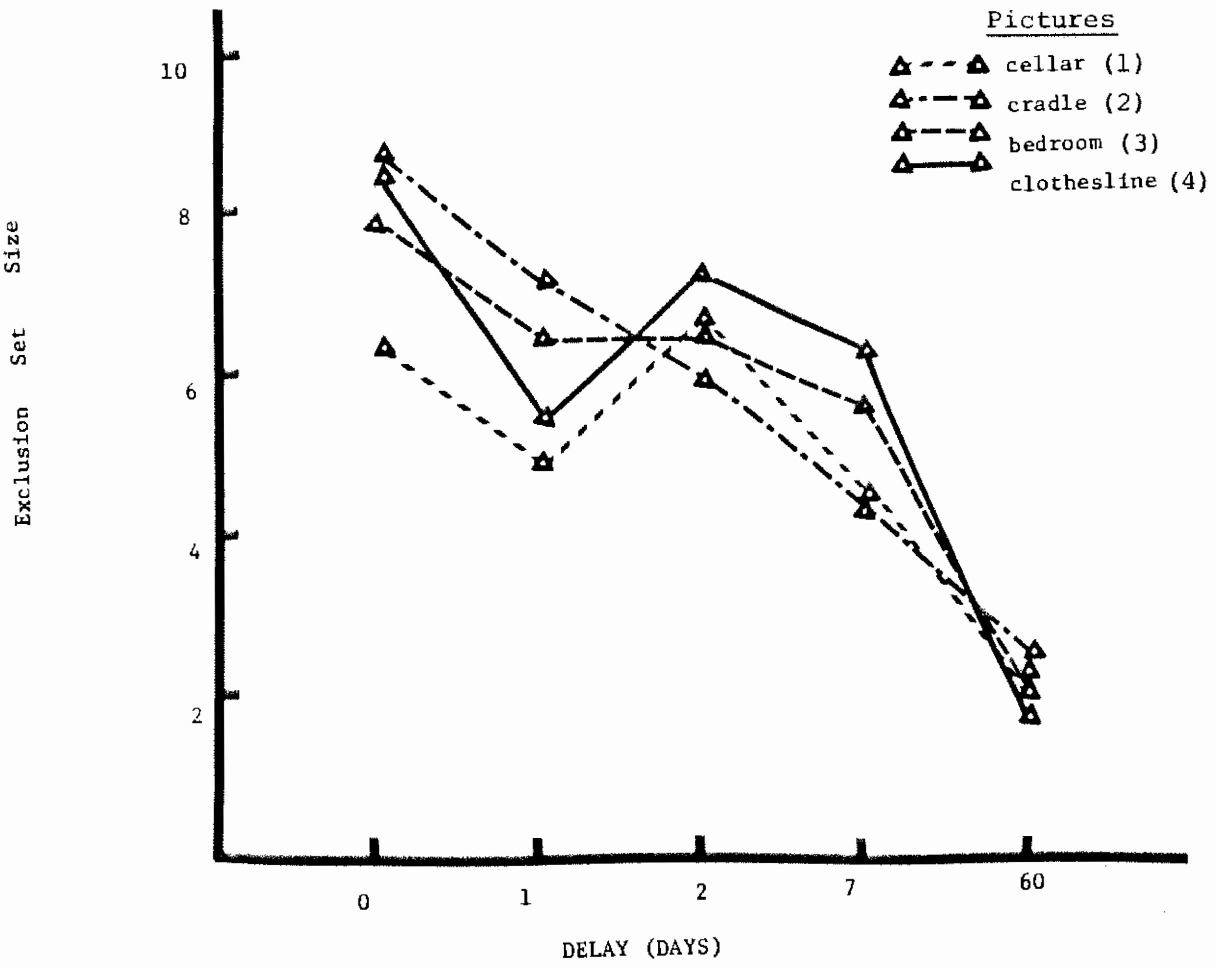


Figure 4. Mean exclusion set scores at each delay period tested for each of the four stimulus pictures (Experiment 1).

differences between means of the delay groups were significantly different. The no-delay group differed significantly from the seven-day and the 60-day group, but not from the one-day and two-day groups. The one-day group and the two-day group differed significantly from the 60-day group but not from each other. The seven-day group differed significantly from the 60-day group. The 10% level of significance was used as the criterion of acceptance, as recommended by Scheffé [1954] due to the conservatism of his test.

Table 3

Summary of the analysis of variance: Experiment 1

Source of Variation	SS	df	MS	F	P
Between Subjects	2797.586	69			
Delay	1043.086	4	260.772	9.661	< .01
Subjects within Groups	1754.500	65	26.992		
Within Subjects	1263.500	210			
Pictures	42.543	3	14.181	2.470	NS
Delay × Pictures	101.171	12	8.431	1.468	NS
Pictures × Subjects within Groups	1119.786	195	5.742		

Table 4

The frequencies of incorrect rejections of the prototype for each delay group in Phase One and Phase Two testing.

Delay Group	Phase 1 Rejection Ratings			Phase 2 Rejection Ratings		
	5	6	7	5	6	7
No-delay	2	9	7	14	6	1
One day	11	2	3	11	4	1
Two days	9	5	4	6	13	1
Seven days	10	4	4	8	2	0
60 days	7	4	1	12	1	0
Total	39	24	19	51	26	3

Table 4 shows the frequencies of incorrect rejections of the four prototype test pictures in all the experimental groups, for Phase One and Phase Two testing. In both phases, subjects rejected the prototype when it appeared with a "5", "6" or "7" rating on 29% of occasions. Most of the rejections were given "5" ratings, though, and only 7% of subjects rejected items in Phase One with a "7" rating. In Phase Two, only 1% of subjects rejected the prototype with a "7" rating, indicating that subjects were cautious in using this rating category. This provides some support for the exclusion set method used here, in which only "7" ratings were used in scoring rejections. Apparently subjects were not guessing when an item was rejected with a "7" rating.

The decrease in "7" ratings of the prototype from Phase One to Phase Two testing shows the advantage of pointing our critical details to subjects. Subjects were hesitant in rejecting an item in each pair shown, even though they had a 50% chance of guessing correctly. To see whether Phase Two testing changed memory scores for subjects, the number of subjects who had received higher exclusion set scores after Phase Two testing than after Phase One testing was counted. Scores were improved by 64 subjects, or 91% of all subjects.

An analysis of subjects' comments in the post-recognition interviews is shown in Table 5. Not all subjects answered the questions in an interpretable manner. About half the subjects in the delayed groups tried to remember the pictures in the interval, but there was no obvious correlation between rehearsal efforts and length of delay. In questioning subjects about their performance on the recognition test, many subjects in each group reported that they were not sure the original was included in the test sets. This reported tendency is supported by data from Table 4, which show that subjects often rejected the prototype. Most subjects also thought that many pictures in each set were the same; they did not notice small differences between the pictures. This appearance of repetitions in the list may have affected subjects' responses by encouraging them to hypothesize about the way the sets were formed. Many subjects (36%) thought repetitions suggested that the picture was more likely to have been the original and six subjects reported distress when a rejected picture was "repeated". A smaller proportion of subjects (11%) thought a repeated item was less

Table 5

Frequencies of responses to questions asked in the post-recognition interview: Experiment 1.

Question	Responses for Each Group				
	No Delay	1 Day	2 Days	7 Days	60 Days
1. Did you try to think of the pictures over the past few days (or weeks)?	NA				
A. No rehearsal		7	4	5	5
B. One rehearsal		2	6	6	2
C. Two or three times		3	3	2	4
D. Several times		1	1	0	3
E. No response		1	0	1	0
2. Was the original picture in each set of pictures in the memory test?					
A. Yes	8	5	5	4	4
B. No	1	0	0	1	0
C. Not always	1	2	3	4	1
D. Not sure	4	7	6	5	9
3. Did you think there were repetitions in the sets?					
A. Yes	13	14	12	12	13
B. No	0	0	0	0	0
C. Unsure	1	0	2	2	1
4. What effect did the repetitions you saw have on your memory?					
A. No effect	1	3	3	1	4
B. Thought it was more likely to be the original and					
(i) this influenced my score	5	3	3	7	4
(ii) this did not affect me	0	1	2	0	0
C. Thought it was less likely to be the original	1	1	2	1	3
D. Upset by repeated items which the subject had rejected	2	0	2	0	2
E. Upset by repetition of an item which the subject had given a positive rating	0	0	0	1	0

likely to have been the original, and one subject said he was upset when an item he had rated positively appeared to have been repeated.

3.2.3 Discussion

The Effect of Delay on Memory for Pictures

The results of this experiment partially supported and extended the findings of earlier studies using homogeneous recognition sets. Memory for pictures declined significantly in accuracy after one week and again after two months of delay. There was no decline over the first 48 hours, a finding which failed to replicate results of an earlier study using homogeneous sets of inkblots and snowflakes as stimuli [Goldstein and Chance, 1970]. It may be that different kinds of pictorial stimuli are forgotten at different rates. Since no recognition method yields an absolute score of memory accuracy, regardless of the demands of a particular paradigm, it is not possible to describe a "true" mathematical function relating memory performance to delay which would apply to every situation. The trend in the function relating exclusion set scores to days of delay included significant linear and non-linear components (found in a trend analysis not reported). However, this should not necessarily be taken as a general property of memory decline, since the exclusion set scores are not equally-spaced points on a linear scale of accuracy of memory. For example, the non-linearity is possibly due to the difference between the small changes of detail made in the lower variations of the test sets and the large changes made in the higher variations.

The present experiment fails to replicate findings of heterogeneous recognition studies, which show that memory capacity for pictures is extremely high and decays slowly over time. The exclusion set recognition paradigm appears to be far more sensitive than a procedure using heterogeneous pictures in measuring memory. In the exclusion set method, a large amount of remembered detail about an individual picture is measured, rather than memory for the class to which the picture belongs (as in heterogeneous recognition studies) or memory for only a small feature of the picture (which is adequate for making a correct recognition decision in a heterogeneous recognition task and many homogeneous recognition tasks in which test pictures differ from each other in many ways). Thus the decline found in memory for pictures over two months in the present study is a unique finding, because partial memory was better controlled than in any previous study investigating effects of delay. It is unlikely that estimates of subjects' average memory scores in each delay group were inflated by subjects' reliance on partial memory, so the decline in memory should have been due to a real loss in the amount of detail in each individual picture which subjects remembered.

The four stimulus pictures were remembered equally well by all groups of subjects, but this should not necessarily be taken as evidence that the decay function found is a "true" function which could be generalized to other situations. The four pictures were chosen because they yielded consistently similar memory scores when pilot subjects were tested in recognition without any delay. This may mean that the underlying scale of memory accuracy is the same for exclusion set scores in all sets of pictures, but it could also mean that variation has been artificially limited by choosing pictures whose differences in difficulty are masked by the method of testing and the choice of distractor items of equal similarity at each level of variation. It is not possible to separate true effects of stimulus difficulty and other effects produced by the testing method itself, such as the similarity of test items, in any recognition paradigm.

It is not clear whether subjects perceive longer delays as more difficult in memory processing than shorter delays. Post-recognition interviews with subjects showed that there were no clearcut differences in rehearsal strategies for different delay groups. Subjects in the longer delay groups did not report thinking about the pictures seen very much more than subjects tested after shorter delays, though they had much more opportunity to do so. Approximately one-third of subjects in each group reported no rehearsal at all, suggesting that pictures are retained in memory without much active concentration or mnemonic strategy. This may be evidence for the view that pictures are remembered "automatically". If this were true, however, the process would not seem to be a very efficient one, as much detail in each picture was misperceived or forgotten even when memory was tested immediately.

There was some evidence that subjects encoded pictures verbally or by some other means in which little visual detail was stored. Three subjects in the 60-day group claimed during testing never to have seen the stimuli, though they admitted having seen pictures with the same types of items. These subjects may have remembered only words which could describe the scene verbally, but this is doubtful since they actually rejected items for "looking wrong". It is more likely that these subjects had some visual detail about the scene in memory which was incorrect. This error in recognition may be analogous to subjects' semantically-based confusions in reproduction (Chapter Two); subjects apparently remembered the classes to which items in the pictures belonged, but had little visual, structural information about them. Memory may become distorted over time so that confabulations are produced in recognizing. In the present case, the stimulus reconstructed in memory may be incorrect even though the subject has some information about the scene portrayed, because:

(1) the subject remembers too little structural detail and cannot use this to reconstruct the whole picture, or

(2) the structural detail the subject remembers is also incorrect.

Little is known about the way information is distorted in memory. Unfortunately, no available recognition method measures memory which is distorted in this way, since the distortions are idiosyncratic and difficult to predict.

The exclusion set method has advantages over other existing methods. For example, 91% of subjects improved their recognition scores from Phase One to Phase Two of testing. This suggests that the ordinary recognition paradigm, where subjects have only one attempt to make a recognition decision about each picture, does not estimate memory capacity adequately because failure to attend to the crucial details in the picture are confounded with memory failures. This is a problem for all recognition tasks. Another advantage of the two-phase testing procedure was that the tendency to reject the prototype item with a rating of "7" was much lower in Phase Two than in Phase One. The final estimate was that only 1% of subjects thought they could positively reject the original picture. This estimate would have been much higher if Phase Two testing had not been instituted. This suggests that in Phase One the original, in the context of so many incorrect distractors, often looked unfamiliar to subjects. Only 37% of subjects were certain that the original had been included in the test sets. On the other hand, when subjects were shown a pair of pictures differing in only one detail, they were often quite able to reject the item which was not the prototype. Of course, in this situation the subject has a 50% chance of guessing the prototype, but this would increase the total number of rejections. Data in Table 4 show that most rejections in Phase Two were "5" ratings, rather than "7" ratings. This shows that subjects do not make extreme rejections unless they are certain about their memories. This gives further justification for the use of only "7" ratings in deriving exclusion set scores: rejections are not counted when subjects are only guessing. Incorrect rejections of the prototype in each set did occur occasionally, but only three subjects rejected the prototype with a "7" rating in Phase Two testing, and hence this was not a major problem with the scoring method.

The fact that incorrect rejections of the prototype occurred fairly frequently with ratings of "5" and "6" give further justification to the use of an exclusion set rather than inclusion set in scoring. Apparently subjects sometimes reject pictures on the basis of incorrect information, so it is necessary to question subjects about their rejections and to create a score based only on correct rejections. In general, subjects seemed to be inconsistent and unpredictable in their positive feelings of familiarity about a picture. This was probably amplified in the present testing situation, where pictures were easy to confuse with each other. Thus it is difficult to specify the subject's "inclusion set" if there is a chance that it will not contain the prototype. For example, subjects who have distorted or confabulated memories will reject all recognition test pictures, including the prototype, because there is none which corresponds to the picture in memory. In this situation, it would be easy to specify an exclusion set score (zero), but difficult to specify an inclusion set. Finally, there was some evidence that exclusion of incorrect stimuli but not inclusion of familiar stimuli is a characteristic of the normal recognition decision:

many subjects spontaneously reported that they were able to exclude pictures which clashed with their memories of the original, but were uncertain about the other pictures.

Problems with the Exclusion Set Method and Suggestions for Its Improvement

Some problems with the exclusion set method were noted. In Phase One testing, subjects did not notice most of the details changed, showing that the changes may not have been discriminable in memory testing. Pilot testing had shown that the differences were perceptible to subjects when they examined pictures simultaneously, but this would have been a much easier task. This meant that Phase Two testing was often lengthy and contributed most to scoring.

A methodological criticism of the paradigm is that all subjects do not see the same number of pictures in recognition testing. Thus the effect of interference on memory scores (subjects may have poorer scores when more test pictures are shown) will not be the same for all subjects. This is an unsatisfactory procedure because subjects did seem to find the sets very confusing. Most subjects thought the recognition sets contained repeated pictures. Thus the pictures were clearly confusable. A more satisfactory procedure would ensure that all subjects see the same number of pictures. The procedure is changed to accommodate this idea in the next two experiments, by showing each subject all pictures in the test sets.

One problem with allowing subjects to view all the pictures in the test sets is that a subject may reject a distractor for the right reason and then fail to reject a higher variation which contains the detail the subject has already rejected. If this occurs because the subject has failed to attend to this crucial detail in the picture, then a procedure such as Phase Two testing, which draws the subject's attention to critical details, can overcome this. On the other hand, the inconsistent memory performance could be due to the fact that the subject has changed the content of what is remembered. The effect of interfering material may be to replace the old, incomplete structural description in memory with a more complete description based on information in the distractor pictures or the subject's confabulations. In this case it may be difficult to define a valid exclusion set score based on a subject's static memory for an item. This issue is discussed again in the later experiments.

Some information about biases in responding was provided by subjects' reports that they guessed which items in each set were correct. Most subjects tended to be more positive about accepting "repeated" items, although some were not influenced and some felt less positive. This also appeared to depend on the subject's initial rating of an item; an item initially rejected might lead to a more positive rating of a similar item which appeared to be a repetition. Unfortunately, there is no way to take into account subjects' biases from the data, since there is no record of which pictures subjects thought were repetitions of others. Subjects were probably overwhelmed by the interference between so many pictures and had a general impression of sameness rather than a distinct listing of particular repetitions. There did not appear to be a general bias influencing all subjects' responses due to the confusing nature of the task. Subjects should be informed of the nature of the recognition sets (i.e., that all pictures are different though they may look the same) so that their memory strategies are more predictable. The interference problem could be lessened by using a smaller distractor set to test each picture.

Another problem is that it is not strictly legitimate to treat the exclusion set scores based on different test sets as additive. This was not a serious problem in the present study, since the effect of delay on memory for each picture was the same. It may be suggested that multidimensional scaling of similarity of items in the test sets could lead to scores which are based on the same scale and hence could be added together to make a general exclusion set score. However, it would be just as difficult to compare interval scales derived from different pictures as to compare the ordinal scales used, since scales are relative only to their components. Thus the combination of different scaled scores into a general score for each picture would never be an absolute measure of memory. It might be a more accurate measure of how much incorrect detail was noticed in the distractors, but on the other hand, the details changed are supposed to be a random sample of all possible details. By using an ordinal rather than an interval scale, the exclusion set recognition measure is not tied to concrete inferences about subjects' tendency to notice detail changes as a function of their size or salience, but rather on the more general notion that cumulative changes will increase the probability of detecting distractors.

Of course, there is no way to determine whether subjects notice changes which clash with their memories on the basis of the cumulative effect of the changes or because high variations on the continuum sample more individual changes likely to be salient for any subject. There is evidence that global changes are important in making recognition decisions, since items more similar to an original picture are more difficult to reject than items less similar, even when all pictures are quite different drawings [e.g., Bahrick and Boucher, 1968]. At the same time, the wider the sample of details presented to a subject, the more likely the subject is to be able to use partial memory to reject an incorrect picture. In the exclusion set paradigm, details changed in any test set are a random sample of details the subject might remember, so partial memory will only aid those few subjects who remember only one or two details which happen to be tested early in the recognition set. Mean exclusion set scores of a group of subjects will be only marginally affected by this distortion of accuracy scores produced by partial memory.

3.3 CONCLUSION

The exclusion set method for testing recognition memory for pictures was used successfully to demonstrate the decline of memory over one week, and again after two months. Since effects of partial memory on recognition scores were minimized, the exclusion set estimate of memory capacity at each delay period tested was lower than in most previous studies. Subjects seldom remembered more than half of all details chosen as critical in a picture and varied in the distractor sets. This suggests that memory for pictures is not in any way "photographic", and that the high estimates of memory capacity in previous studies may have been due to the fact that subjects were given credit for remembering a picture when in fact they did not remember many critical features of the stimulus. The exclusion set method does not yield scores which are absolute measures of memory, but it does provide a sensitive test of detail remembered in a single picture which can be used to compare subjects' memories in different experimental conditions. It is potentially very useful in testing memory for pictures in a variety of situations, since it yields an accuracy score for each stimulus tested, based on several responses from the subject concerning the stimulus, rather than a single "old" or "new" recognition response. Hence the method does not require a stimulus set of hundreds of pictures, which would be difficult to control, such as has been used in conventional paradigms. The greater number of observations recorded about a single stimulus in the exclusion set paradigm allows a smaller set of stimuli to be used.

CHAPTER FOUR

INTENTIONAL AND INCIDENTAL INSTRUCTION IN RECOGNITION MEMORY FOR PICTURES

The findings of the experiment reported in Chapter Three, in which a sensitive homogeneous recognition task was employed to measure the decay in memory over time, suggest that memory for pictures has a limited capacity, at least when several items are presented for short periods of time. Few of the pictorial stimuli were totally forgotten, but subjects differed widely in the amount of detail they remembered about the pictures. This suggests that memory for pictures is not "quasi-photographic" and that it can be increased or decreased by choosing appropriate experimental conditions.

Since memory for pictures does not appear to be limitless, the next question that must be asked is whether encoding of pictures is automatic. If pictures are remembered just as well when they are given a quick glance as when they are carefully observed, then encoding may reach its stage of maximum processing in a very short time for all pictures. The studies which have demonstrated high capacity in memory for pictures have been taken to imply that recognition of pictures is an automatic process as well as a quasi-photographic one. Experiment 2 was carried out to examine the complexity of the encoding process, to see whether it is an active, variable operation.

If encoding of pictures can be manipulated so that one kind of encoding strategy leads to more accurate memory than another less efficient strategy, then it is possible to conclude that encoding is not automatic and is an active process influenced by the demands of the situation. Some evidence for this view is provided by studies which show that memory for pictures is influenced by the appropriateness of instructions which inform subjects whether the task is to be recognition or recall [Tversky, 1973, 1974]. This suggests that subjects must attend to relevant details in the pictures in order to remember them and that encoding may not be adequate when subjects are not given appropriate instructions. The manipulation of instructions to subjects provides a technique for controlling encoding strategies. An instructional manipulation traditional in verbal learning studies has been the subject's intention to learn in the memory task. In Experiment 2 subjects' encoding strategies were manipulated by giving groups of subjects different instructions so that some subjects expected the memory task while others did not. It was hoped that this might induce two maximally different encoding strategies and hence provided the most favourable conditions in which encoding effects could be demonstrated.

4.1 EFFECT OF INSTRUCTIONS ON RECOGNITION AND RECALL

4.1.1 Studies of Intentional and Incidental Memory

It was mentioned in Chapter Two that most researchers argue that recognition is a fairly simple process in which a familiarity decision is made about a test item, while recall requires a search of storage and more elaborate retrieval of information from memory [Kintsch, 1970; McCormack, 1972]. This view is based largely on studies showing that recognition memory is generally higher than recall for the same verbal material. Recently researchers have put forward evidence refuting this view, by showing that recall of words can exceed their recognition in certain circumstances [e.g., Watkins, 1974]. The exclusion set method is a recognition paradigm which certainly seems to require that subjects retrieve quite detailed information about a pictorial memory trace, suggesting that a recognition task can be as complex as a recall task.

Results of studies which show that intentional and incidental instructions have different effects on recognition and recall of words have been used to support the hypothesis that recognition is a simpler process than recall. It is argued that if recognition and recall are measuring the same memory process, then they should be affected equally by instructions which manipulate how much information subjects have about the memory task. A common finding [e.g., Eagle and Leiter, 1964] has been that free recall of words is adversely affected if subjects are not told about the memory task (i.e., when subjects are given "incidental instructions"), but that recognition of words is the same whether subjects are given incidental instructions or intentional instructions. In the latter case, subjects are informed about the memory task. Since subjects recognize words whether they are trying to remember them or not, it might appear as though recognition is a fairly automatic process, while recall requires elaborate encoding and rehearsal strategies and a concerted effort at retrieval during testing. Even if this were true for words, it need not necessarily be true for pictures. Also, it is difficult to compare scores of memory accuracy based on recognition or recall methods, for reasons outlined earlier. However, it is possible to test recognition memory under intentional and incidental instructions, to see whether subjects can encode more information from pictures when they are aware of the memory task. Experiment 2 examines this question.

There is some evidence that incidental memory can be high or low in accuracy depending on variables in the test situation. For example, in recall, when meaningless items are used incidental memory does not differ as much from intentional memory as when meaningful material is used [Mechanic, 1962; Postman, Adams and Phillips, 1955]. Similarly, the difference is reduced when fewer cues are given in the recall task [Postman *et al.*, 1955], or when the words in the list are less related to each other [Postman and Adams, 1957]. Incidental recognition has actually been found to be superior to intentional recognition when pictures are presented for 0.25 seconds, but instructions have no effect on recognition when pictures are presented for four seconds or recall when pictures are presented for 0.25 seconds [Dornbush and Winnick, 1967]. These findings suggest that incidental memory, much like recognition memory in general, is a function of many variables in the testing situation.

4.1.2 The Orienting Task

In incidental memory groups, subjects are usually given some task to perform with the stimuli: an "orienting task". This is given so that subjects will be occupied in a meaningful way during presentation of the stimuli so that they will not anticipate the real nature of the task. The orienting task is also given so that subjects will actually attend to each item, so that the difference between intentional and incidental subjects is in their motivation to remember, rather than in viewing time or attention paid to the stimuli. Orienting tasks may help subjects to remember the stimuli, as for example, when the task is to classify each item in the list which is later to be recalled, or they may hinder remembering, for example, when the task is to name the colours in which the words are printed. Since memory performance under incidental conditions can be made very accurate by varying the orienting task conditions or stimulus materials, McLaughlin [1965, p.274] concluded that:

> "There is no experimental evidence demonstrating (incidental) learning in the strict, traditional sense — namely, a distinct learning process which occurs when there is no motive, self-instruction, or set to learn."

Postman [1964] also agreed that there is no reason to postulate conceptually different processes in intentional and incidental memory, since an appropriate orienting task could facilitate memory to the same extent as intentional instructions.

The discovery that different orienting tasks may affect memory in different ways led to research into "levels of processing", which assumes that:

> "... memory performance is a positive function of the level of processing required by the orienting task" [Craik and Lockhart, 1972, p.678].

This view does not suggest that the difference between intentional and incidental memory performance is due to retrieval differences. Rather, different instructions are thought to produce differences in encoding processes, and it is hypothesized that some orienting tasks may aid subjects' encoding of stimuli, while others lead to poor encoding. It is actually difficult to determine whether differences between any memory groups are due to encoding or retrieval processes [see Lockhart *et al.*, in press].

Evidence for differential effects of various orienting tasks has come from verbal learning experiments which show that an incidental orienting task requiring subjects to respond to words by giving semantically-based ratings (e.g., pleasantness) leads to recall performance equal to that of intentional subjects, while subjects who perform a lexical task of examining words for the presence of a certain letter or estimating the number of letters recall fewer words [e.g., Hyde and Jenkins, 1969]. Hyde and Jenkins [1969] found no effect on the difference between intentional and incidental memory performance of increasing presentation time from two to four seconds or repeating stimuli a second time. This contrasted with some of the earlier work and suggests that orienting tasks, not motivation *per se*, contribute to instructional differences. The superiority of incidental, semantic tasks over "structural" or non-semantic tasks has also been found in recognition of words [Schulman, 1971], recall of sentences [Rosenberg, Schiller and Smith, 1974], and cued recall of both related and unrelated word pairs [Epstein, Phillips and Johnson, 1975]. This attests to the usefulness of specifying orienting tasks to predict memory performance.

One experiment has found differential effects of orienting tasks on recall and recognition. Griffith [1975] asked subjects to perform one of two orienting tasks: (1) rating words for their imagery value, or (2) categorizing them into conceptual classes. Afterwards, the subjects were required to recall or recognize the words. Imagery rating led to better performance in recognition than word categorization, while the categorization task led to superior recall. Griffith concluded that imagery elaboration is more facilitative to recognition processing, while organization is more facilitative to recall. This finding seems to fit with the notion that recognition is based on an automatic retrieval which does not require elaborate organization of information remembered, while recall is highly elaborated and structured. However, studies by Sheehan [1972, 1973] suggest that imagery can aid memory in intentional situations but not when incidental instructions are used, no matter what paradigm is used for testing retention. Memory for abstract nouns was found to be as high as memory for concrete nouns with intentional instructions, but was much lower with incidental instructions and an orienting task of making familiarity judgments about the words. This was demonstrated for both recognition and recall. Sheehan suggested that "imagery mediation" aids memory, and, since concrete nouns appear to be better than abstract nouns in evoking images, they are better remembered in situations when subjects are not trying to commit the stimulus to memory. On the other hand, abstract nouns are not usually image-evoking and hence would not be remembered with the help of images unless subjects were employing some intentional strategy which increases their tendency to image to a word. Sheehan's work suggests that imagery can be used as an intentional strategy to aid both recognition and recall. Perhaps the imagery instructions given to Griffith's [1975] subjects were not appropriate to the task. Thus it is difficult to specify the effects of various orienting tasks on encoding strategies to determine which will be most facilitative in any testing situation.

4.1.3 Instructional Effects on Memory for Pictures

Few studies have used pictorial stimuli in varying remembering instructions experimentally. Paivio [1974] examined verbal recall of pictures under incidental and intentional instructions in order to study effects of spacing of repetitions in memory. For both pictures and words, presented either once or twice in a list, a free-strategy intentional group scored higher than a group given intentional instructions together with an orienting task of guessing whether the next item in the list would be a picture or a word, while an incidental group given the orienting task only performed most poorly. This study replicates findings of studies using words as stimuli, but this may have been due to the use of written verbal recall as the response for remembering the pictures.

One recognition study using pictures has included groups given incidental and intentional instructions. Bower and Karlin [1974] compared recognition of faces with incidental or intentional instructions using an orienting task in which faces were rated for their honesty, likableness or the sex of the person portrayed. They found no differences in hit rates or confidence ratings across the different intentional and incidental groups, though within each group there were significant differences between the orienting tasks. Judgments about the sex of the faces led to consistently poorer recognition than judgments about their honesty or likableness. Judgments of sex were assumed to have been made on the basis of physical characteristics, while the other judgments may have required more semantic analysis. This use of three different orienting tasks is an attempt to replicate the design of many depth-of-processing experiments with nonverbal material. The problem with Bower and Karlin's [1974] study is that it is difficult to predict the depth of analysis in subjects' processing by manipulating experimental conditions. For example, the subjects in Bower and Karlin's experiment saw each face for five seconds, which was probably more than enough time to make a judgment about the sex of the person portrayed in the picture. What were those subjects doing in the remaining time? The results suggest that perhaps sex-judgment subjects were not attending as much to detail in the stimuli as subjects making seemingly more complex judgments and may have been processing them at a more shallow level. On the other hand, it may be assuming too much to suggest that subjects attending to physical features of a picture do not encode the whole picture for its meaning. Whereas it appears intuitively possible to proofread sentences without being aware of their meaning, it seems implausible to suggest that we can process pictures at a surface level without understanding what the pictures represent.

4.1.4 The Present Approach

The first step in determining whether recognition of pictures is affected by different strategies is to see whether incidental and intentional instructions have any effect on recognition of pictures. It has been seen that instructions affect recall of pictures and words, but not necessarily recognition of words. No study has used pictures other than faces to test these differences in recognition. Furthermore, evidence presented in Chapter Two showed that standard recognition procedures yield insensitive measures because of ceiling effects produced in distributions of accuracy scores. The exclusion set method appears to be a sensitive measure of recognition accuracy which might expose subtle differences between groups not found in other studies. The aim of the experiment reported in this chapter was to compare recognition performance for subjects given remembering instructions with performance of subjects given an orienting task without any remembering instructions. A Control group was given the orienting task with remembering instructions. It was predicted that the Intentional group given no orienting task would perform best, as suggested by previous recall studies, since the orienting task would not be as facilitative as the subject's freely-chosen strategies. The relevant comparison in this case would be between the Intentional group and the Control group. To test effects of motivation to remember on recognition performance, the Control group was compared to the Incidental group. Since both groups were given an orienting task, and differed only in whether they had been told of the memory task, it was hypothesized that if motivation is important in recognition of pictures, then the Control group would have superior recognition. The results of Bower and Karlin's [1974] study suggest the opposite hypothesis, that motivation *per se* has no effect on recognition of pictures.

In Experiment 2, the exclusion set method was changed to control the interference effects which appeared to be biasing memory accuracy scores in Experiment 1. Phase One testing was omitted and Phase Two testing was standardized so that each subject saw 15 pairs of test pictures for each stimulus in recognition. This ensured that all subjects saw the same amount of material, which appears to be a necessary control if interference between pictures is to be held constant for all subjects.

4.2 EXPERIMENT 2: THE MANIPULATION OF INSTRUCTIONS

4.2.1 Method

Subjects

To fulfil a course requirement, 45 first-year Psychology students participated in the experiment. There were five males and 10 females in each of the three experimental groups.

Stimuli and Apparatus

These were the same as those described in Experiment 1.

Design

There were three experimental groups. The Intentional group was given no task to perform while viewing the stimuli; subjects were told to try to remember as much as possible about each picture, using any remembering strategy they chose. Subjects in the Control and Incidental groups were given an orienting task of counting repetitions of classes of items in each picture (e.g., dishes), where subjects were allowed to define the items in each picture according to any classification system they chose. Control subjects were told that a recognition task would follow the presentation of the pictures, while Incidental subjects were not. The orienting task was chosen because it allowed subjects to classify items in the pictures into semantic categories without forcing them to search for a particular class of items, which seemed to be too easy a task, given the stimuli available. The task also seemed to be a good "blind" for the Incidental subjects viewing the pictures.

Procedure

Subjects were given one of three types of instructions before testing of recognition, which are described below.

(1) Incidental: Subjects in this group were told by the experimenter,

> "This is a study about how people classify objects in complicated pictures. I am trying to decide which pictures are easiest to extract maximum information from. I want you to notice whether there are repetitions of items in the pictures. This depends, of course, on how you happen to be classifying the items you see. For each picture I want you to say how many kinds of items are repeated. As an example, I'll show you a picture for 10 seconds."

The first practice item (Appendix 1) was then shown for approximately 10 seconds while the experimenter said,

> "You might see that there is more than one tap and more than one fold in the towel, so you would call out 'two' because there are repetitions of two items. Of course, you might not have found any or you might have found several."

A second practice item was shown for 10 seconds, after the experimenter had told subjects to say aloud the number of types of items which the subject thought were repeated in the picture.

(2) Control: Subjects were given the same procedure as in the first group, except that they were first told,

> "I have some pictures I want you to remember so that I can test your memory for them afterwards."

(3) Intentional: Subjects were given the same initial instructions as subjects in the second group and then told,

> "I want you to take in the whole picture first and then focus on each detail. Think about what each item is and note the positions of items, their size and type. Look out for any distinguishing features that might help you remember things. Name things in the pictures when it might help your memory."

The practice picture was presented for 10 seconds, while the experimenter reminded subjects to notice details in the pictures and to name important features to themselves. The second practice item was then shown for 10 seconds.

Subjects in the first and second groups did not receive verbal instruction while the practice picture was being presented. Pilot testing had shown that it is difficult to encourage intentional subjects to focus on the picture in detail without exhorting them to be attentive during its presentation. After it was determined that subjects knew what to do, they were shown the second practice item and no verbalization was given while the picture was presented for any group. Subjects were asked to use their particular attentional strategy with the picture. Thus subjects in all three groups saw two practice pictures, but were not tested for memory of them.

After the initial instructions, subjects were presented with the four stimulus pictures on the right side of the screen in an individually randomized order. Pictures were presented for 10 seconds each, with approximately five seconds interstimulus interval. Afterwards, subjects were told by the experimenter,

> "I'm going to show you pictures two at a time. They will be different in some way, which I'll point out to you, and I want to see if this helps you reject a picture on the basis of this difference. It might jolt your memory. However, you might reject both of them. Don't guess if you are not sure."

Recognition testing differed from the testing carried out in Experiment 1. Only Phase Two testing was carried out. All pictures in the stimulus set were shown in an individually randomized order. Each picture was presented with its neighbour above or below it in the stimulus set, so that the pictures in the pair differed in only one feature. The positioning of each picture in the pair on the right or the left was determined randomly. As each pair in a particular set of pictures was shown to subjects, the experimenter pointed to the difference between the two pictures and asked subjects if they could reject one picture on the basis of the difference. If subjects said they were not sure which picture to reject, they gave a rating of "1" to the pair. Subjects who rejected a picture were asked to give the rejected picture one of the following ratings:

(2) I don't think it could be the original.

(3) I am fairly sure it is not the original.

(4) I am positive it is not the original.

There was also provision for subjects to say that both pictures looked unfamiliar. This was noted by the experimenter, but not counted as a rejection of either picture. Each stimulus was tested separately, with ordering of sets randomized for each subject. Thus subjects saw 15 pairs of pictures for each of the four test sets. After recognition testing, subjects were asked to fill in a questionnaire on the experiment (Appendix 2). One subject in the Incidental group had anticipated the memory task and was rejected.

4.2.2 Results

Analysis of Exclusion Set Scores

For comparison with the experiment reported in the last chapter, the exclusion set sizes for the three groups were scored by subtracting from 16 the number of the lowest distractor variation each subject could exclude correctly with a "4" rating. Means and standard deviations of these are shown in Table 6. The Intentional group was comparable to immediate recognition scores of subjects in the first

Table 6

Means and standard deviations of exclusion set scores for each group: Experiment 2.

	Group		
	Incidental	Control	Intentional
$\bar{X}$	4.967	5.800	6.717
S.D.	2.774	3.167	3.157

experiment. The slightly lower scores in this experiment were probably due to the lack of a practice trial which had been included in Experiment 1.

It was decided that the method of testing in this experiment changed the nature of the recognition memory measured. Since the Phase One testing of Experiment 1 was omitted, there was no measure of global aspects of memory for the stimuli and only memory for the 15 specific detail changes was actually tested. This made superfluous the assumption that successive changes to the stimuli would increase the probability of subjects' rejection of a stimulus as incompatible with their memory representations. A more conventional scoring procedure was used in the recognition accuracy analysis in order to compare the groups.

Analysis of Accuracy Using Confidence Ranges: Hit Rate Analysis

Table 7 shows the means and standard deviations of distributions of confidence ratings of incorrect responses for each group. All rejection responses were used in determining each subject's score, since it appeared that groups may have differed only in their propensity to use ratings of "1".

An accuracy score was derived by subtracting all incorrect rejections from all correct rejections and dividing these by the total number of non-guessing responses. Ratings of "2", "3" and "4" were pooled in this analysis and given equal weighting. Guess responses were not included in the analysis because there appeared to be an effect of

Table 7

Means and standard deviations of the frequencies of correct and incorrect responses for particular confidence ratings: Experiment 2.

Group	Confidence Ratings						
	Guess Responses	Correct Responses			Incorrect Responses		
	1	2	3	4	2	3	4
Incidental							
$\bar{X}$	18.200	6.733	8.000	11.800	7.333	5.867	2.200
S.D.	9.360	2.865	3.423	4.329	2.992	3.720	2.077
Control							
$\bar{X}$	15.133	8.133	9.867	12.000	7.467	5.467	2.067
S.D.	7.434	3.796	4.969	6.730	3.378	3.623	1.751
Intentional							
$\bar{X}$	13.400	6.800	8.467	16.067	5.333	6.667	3.267
S.D.	6.555	3.052	3.335	6.397	2.820	2.895	5.365

the experimental conditions on the tendency to use "1" ratings. Means and standard deviations of accuracy scores for each group and each picture are shown in Table 8. A planned comparisons analysis was carried out on the accuracy scores in different groups pooled over pictures. In the first contrast, the Incidental group was compared with both intentional groups combined to test the hypothesis that intentional instructions would aid recognition, but no significant differences were found ($F = 1.519$, $df = 1,42$, $p > .05$). In the second contrast, the Control and Intentional groups were compared to test the hypothesis that the orienting task would interfere with performance, but this was again not significant ($F = 0.015$, $df = 1,42$, $p > .05$).

To examine the effect of the different pictures on accuracy scores, a two-way analysis of variance with repeated measures on Pictures as a random variable was carried out on the data. The summary table is shown in Table 9. There was no significant effect of Instructions, but there was a significant effect of Pictures and no significant interaction. Thus, though the pictures differed widely in their memorability, none were affected by instructional variations. There was still a significant effect of Pictures when a second two-way analysis of variance was carried out on pooled accuracy scores for Pictures 1, 2 and 4 only, omitting the badly-remembered third picture.

Table 8

Means and standard deviations of hit rates for each picture in each instructional condition: Experiment 2.

Pictures		Group		
		Incidental	Control	Intentional
1.	$\bar{X}$	.3184	.3133	.3645
	S.D.	.2322	.3028	.2704
2.	$\bar{X}$	.4837	.4758	.4728
	S.D.	.2429	.1881	.2435
3.	$\bar{X}$	-.0095	.0691	.0986
	S.D.	.3750	.3121	.2814
4.	$\bar{X}$	.3604	.5404	.4754
	S.D.	.2343	.2606	.2506
Total	$\bar{X}$	.2883	.3497	.3528
	S.D.	.3271	.3209	.2985

Table 9

Summary of the analysis of variance: Experiment 2.

Source of Variation	SS	df	MS	F	p
Between subjects	5.3144	44			
Instructions	0.1126	2	0.0563	0.4544	NS
Subjects within groups	5.2018	42	0.1239		
Within subjects	12.1909	135			
Pictures	4.8373	3	1.6124	28.5887	$< .01$
Instructions × Pictures	0.2498	6	0.0416	0.7376	NS
Pictures × Subjects within groups	7.1038	126	0.0564		

Signal Detection Analysis

Because of the large guessing scores and the fact that the data analysis was insensitive, a signal detection analysis was carried out on the data to attempt to use more information about confidence ratings as related to accuracy scores. In the previous analysis, no greater weighting was given a correct rejection with a "4" rating than a "2" rating. Many rejections given a "2" rating may be guesses. Signal detection analysis allows more sensitive accuracy and bias measurements, taking subjects' guessing strategies into account.

Following a procedure for deriving nonparametric accuracy, P(A), and bias scores outlined in McNicol [1972], a confusion matrix was produced for each subject, pooled over pictures to increase cell frequencies. The paradigm used applied to a two-alternative forced-choice recognition paradigm with confidence ratings [McNicol, 1972, ch. 5). It was assumed that subjects made both a signal and a noise judgment for each pair of pictures shown during testing; by rejecting a distractor (noise) they implicitly accepted the correct pair member (signal). Figure 5 illustrates the type of confusion matrix used. It was further assumed that there were signal and noise distributions for all items appearing on the left side of the screen during testing, and complementary signal and noise distributions for items on the right side of the screen. This allowed a correction of any bias in subjects' responses towards increased accuracy for left or right items. It is easiest to represent the familiarity of a left or right item by describing it in terms of its magnitude as a signal compared with its magnitude as noise. To choose whether the left or right item is more familiar for recognition, the subject must compare the distribution of familiarity for a left item with the familiarity distribution of a right item. This situation is illustrated in Figure 6. The "A" distribution is actually the noise distribution for left items minus its signal distribution, to create a resultant "certainty of noise" distribution. Distribution "B"

	Ratings for Left Pictures				Ratings for Right Pictures			
	4	3	2	1	2	3	4	
Noise on Left								Signal on Right
Signal on Left								Noise on Right

Figure 5. The confusion matrix.

describes the certainty of noise for items on the right side. (In this experiment, judgments pertained to identifying noise, rather than signals.) Subjects who use "1" ratings of uncertainty with either right or left items are assumed to receive equal or indiscriminable amounts of noise from both items in the test pair; i.e., distributions A and B are nearly equal in area over this interval.

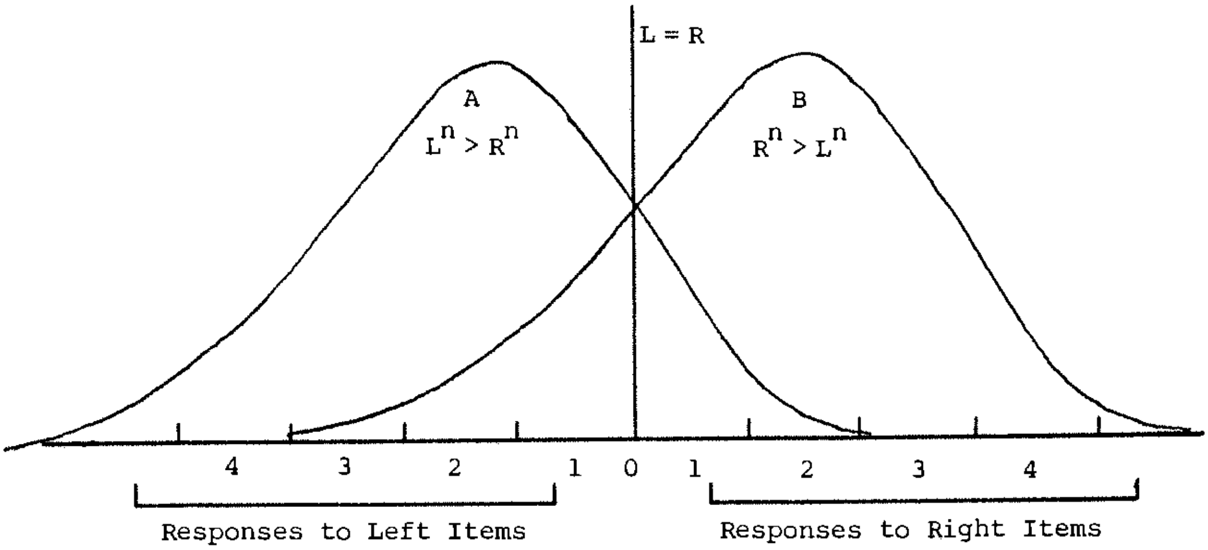


Figure 6. Noise-signal distributions for left items (A) and right items (B). Confidence ratings along the abscissa demonstrate a theoretical subject's responses based on the proportion of noise to signal for either a right or left item. For example, subjects who reject the left item with a "3" rating do so because there is more area of distribution A in this interval than distribution B; i.e., the noise value of the left item (L^n) is greater than the noise value of the right item (R^n).

For the present paradigm, confidence ratings for left and right items were separated to make a seven-point scale. Frequencies of use of these ratings were tabulated for the two possible situations:

(1) when noise is on the left and signal on the right, as in the first row of the matrix in Figure 5, or

(2) when the noise is on the right and signal is on the left, as in the second row of Figure 5.

P(A) and bias scores were calculated for each individual on the basis of frequencies of responses in each cell of these confusion matrices. For some subjects, matrices had to be collapsed to two-by-six, two-by-five or two-by-four matrices to eliminate cells without responses, which distort the cumulative probabilities found for each matrix. A planned comparisons procedure was carried out on the arcsin P(A) scores. This is the usual normalizing transformation in signal detection estimates. Mean arcsin P(A) and bias scores are shown in Table 10. The Incidental group did not differ from the intentional groups combined ($F = 2.293$, $df = 1,42$, $p > .05$) and the Intentional and Control groups did not differ significantly from each other ($F = 0.076$, $df = 1,42$, $p > .05$). In all groups there was a tendency for subjects to make low ratings. A bias score of 3.5 shows subjects' tendency to use the "1" rating category most and to use higher ratings less frequently.

Table 10

Mean arcsin P(A) and bias scores for all groups in Experiment 2.

		Group		
		Incidental	Control	Intentional
Arcsin P(A)	$\bar{X}$	1.8905	1.9660	1.9837
	S.D.	0.1625	0.2028	0.1600
Bias	$\bar{X}$	3.4959	3.5544	3.5064
	S.D.	0.1960	0.3407	0.2585

The signal detection analysis was more powerful than the hit rate analysis because it eliminated the response bias which differed according to the different groups. Hence the greater use of "1" ratings in the Incidental group was corrected by the analysis. However, there were still no differences in performance under the three instructional conditions.

Questionnaires

The post-test questionnaire was used to determine whether subjects in each group had tried to remember the pictures or not; i.e., whether they had conformed to experimental instructions. To determine

Table 11

Frequencies of responses to questions about intention to remember and rehearsal in Experiment 2.

Question		Group		
		Incidental	Control	Intentional
1. Was the memory task expected?	Yes	1*	15	15
	No	15	0	0
2. Did you try to remember the pictures during presentation?	Yes	1*	5	15
	No	15	10	0
3. Rehearsal:				
None		10	10	2
Reviewed orienting task responses		1	1	0
Imaged		1	0	7
Verbalized		0	0	1
Unspecified rehearsal		3	4	5

* This subject was rejected and is not included with subjects answering Question 3.

whether they were rehearsing, subjects were also asked what they did in the interstimulus intervals. Table 11 shows the results of the analysis of protocols. It is interesting to note that it was possible to control subjects' expectations about the task but that this did not automatically control subjects' memory strategies. Most of the control subjects reported that they were unable to "try to remember" the stimuli; i.e., that the orienting task was so demanding that they forgot about remembering the items during presentation. It can also be noted that the pattern of rehearsal was markedly similar for the Incidental and Control groups; in both groups 73% of subjects did not try to rehearse the items. In the Intentional group 87% of subjects attempted some sort of rehearsal.

4.2.3 Discussion

The Effect of Instructions to Remember on Recognition of Pictures

The findings of the present experiment concerning recognition of pictures are in keeping with findings for recognition of words [Eagle and Leiter, 1964; Postman *et al.*, 1955], which have been used to suggest that recognition is a fairly simple process unaffected by elaborate encoding or retrieval strategies. The finding that there were no obvious intentional-incidental differences for any of the four pictures used, despite the heterogeneity of their content, suggests that the results were not specific to a narrow range of pictorial materials. Motivation to remember does not appear to aid recognition of pictures.

There were lower recognition scores in the present study than in Experiment 1, where subjects were given two practice trials (after scores were made comparable). This may mean that the intentional instructions used in the present study did not prepare subjects for the elaborate attentional and/or retrieval strategies that they could deduce for themselves in a practice trial. This difference suggests that the pictorial recognition process is an elaborate one, and does require attention to relevant stimulus details. In this case, a practice trial would be a better operational definition of intention to learn than listening to the experimenter's exhortations to subjects about remembering, which are often misunderstood. Unfortunately, the practice trial would also give the intentional group an unfair advantage in, for example, seeing the material, while the incidental groups obviously could not be given a practice trial. In Experiment 3 (Chapter Five) a group given practice with the task was compared with subjects given orienting instructions to test this idea. There was an attempt to ensure that subjects in all groups received similar exposure to the task and materials.

The orienting task used in the present experiment, noting repetitions of objects portrayed in each picture, may have been as facilitative in guiding subjects' attention to details in the pictures as the vague instructions given to Intentional subjects. Evidence for this comes from the finding that two pictures were recognized best (on average) by the Control group. A similar trend was also evident in pilot testing. However, protocol analysis revealed that control subjects found it difficult to try to remember the pictures and perform the orienting task at the same time. They also reported rehearsal of the items between picture presentations far less than Intentional subjects. Thus, defining orienting tasks is not merely a matter of identifying how facilitatory or interfering they are [Postman, 1964]. In any particular task the subject brings to the situation old habits with regard to memory strategies and these may not be appropriate to the task. The subject must suppress these and replace them with the experimenter's intentional instructions. However, in the situation in which intentional instructions are given with an orienting task, the subject may try to combine the old strategies with the task required by the orienting instructions in an attention-sharing process. Hence even a facilitating orienting task may make the remembering difficult, unless subjects are told how to *use* the orienting task as their strategy in the situation. Instructions to subjects must be carefully defined, as the same orienting task may be perceived by subjects in different ways. In this experiment, both the freely-chosen intentional strategies and the orienting task instructions may have been inappropriate for adequate memory encoding, and may have failed to test the effects of beneficial encoding strategies on memory accuracy.

While the instructions did not have a marked effect on recognition accuracy, they did affect confidence in memory judgments. The tendency to be positive in confidence ratings was much greater for the Intentional group than the groups given an orienting task, even though the accuracy of these judgments for the three groups did not differ. The tendency to say "not sure" instead of making a recognition response was greatest for the Incidental group, lowest for the Intentional group, and intermediate for the Control group. Perhaps the results show that the Incidental group did remember less, but that where a rejection response was given, it was equally likely to be correct for all groups. In this situation, it would have been better to force Incidental responders to be more confident by omitting the "no response" category. This raises an interesting problem in these studies. Instructional variations may affect memory scores not through memory accuracy, but through response biases where subjects change their response output criteria according to variations in the experimental situation.

Evaluation of Signal Detection Measures

Many researchers in memory are becoming dissatisfied with signal detection measures in representing accuracy scores. Lockhart and Murdock [1970] pointed out that use of signal detection theory burdens the memory researcher with a number of assumptions about memory processing: constancy of noise factors, the continuous unidimensional sensory continuum based on strength of memory traces, and identification of a unidimensional decision axis when subjects are tested for their memories under various conditions. Recently Marken and Sandusky [1974] showed that the assumptions of signal detection theory were not valid for a verbal recognition task; subjects did not bias their responding in order to optimize their correct responses, and responses were not independent of each other. This would seem to hold true in the exclusion set method, where subjects accept or reject a picture on its second presentation, after making a previous contrary decision because they have attended to inappropriate details. Also, there was evidence of interference among items in each set. Subjects appeared to have an evolving memory trace based on guesses about features in each distractor, especially when they were uncertain about their original memories. This would contradict the assumption that a clearcut decision axis and memory criterion were in use. It is also feasible to suggest that subjects' strategies and hypotheses about optimizing their performance varied over time. Subjects in the different instructional groups appeared to have different strategies based on their criteria for recognition decisions, yet the nonparametric signal detection method used in scoring did not find any differences. The ordinary hit rate analysis appeared to be sufficient for this study. Since signal detection analyses seldom differ from traditional analyses in their findings, it seems prudent to avoid assumptions about memory inherent in the signal detection scoring method and to use instead a traditional analysis based more on knowledge about each individual experiment and its particular innovations.

Problems with the Method for Testing Recognition

While the present study used the same stimuli and apparatus as in Experiment 1, the procedure was changed so that exclusion set scores could not be calculated. The testing procedure in Experiment 1 was unsatisfactory because there were differences in the number of pictures subjects were shown during testing. Since interference effects seem to be important in the exclusion set method, the variation in amount of interference to which different subjects were exposed provided a large source of uncontrolled error in Experiment 1 that was overcome in Experiment 2.

The testing procedure in the present experiment was like a forced-choice pairs paradigm, except that subjects made choices between details in a pair of pictures, not between whole pictures. Also, 15 correct recognition decisions about pairs of pictures were needed to achieve a perfect score rather than a single rejection of a picture in one pair. This method gave a sensitive measure of the number of details sampled in a picture which the subject could reject as unfamiliar, and the number of details (i.e., members of a pair) which subjects could reject can be described as an exclusion set score. There is a problem with this method, however: subjects can reject one of the pictures in each pair without any memory for the stimulus and be correct on 50% of these occasions. Thus guessing strategies are as serious a problem in this method as in the forced-choice pairs method. Since subjects do not have to give a reason for their rejection of a picture (the experimenter points out the difference to subjects in order to prevent attentional failures), there is no means by which guessing can be minimized. It was originally intended in this study to use only extreme ratings of rejection as the criterion for exclusion. However, it appeared that this criterion might be too stringent, and would fail to give cautious responders credit for their accurate memories. Thus, response biases would have interfered with accuracy scores. Since it is likely that subjects in different instructional groups also differ in their response biases, this problem seems to be a serious consideration. For these reasons, the Phase Two recognition testing method used in the present experiment was unsatisfactory in producing exclusion set scores which would reflect real differences in memory accuracy. Further improvements in the method of testing recognition with the exclusion set paradigm were carried out in Experiment 3.

The differences between pictures found in this study demonstrate that the failure to find differences in Experiment 1 (using the same pictures) was not due to the similarity of the pictures, but to a similarity in the method for testing memory for the pictures. Apparently the test sets for the four pictures do not contain detail changes equal in salience for subjects. The testing method in the present experiment allowed an examination of the frequencies of rejections for each member of adjacent pairs in the test sets. There were three details changed in Picture 2 (see Figure 2) and two details changed in Picture 4 (see Appendix 1) which no subjects rejected. Hence these were not differences discriminable in memory. In an ideal exclusion set test, all details changed would be equal in salience for subjects, but this is very difficult to achieve, since salience is a difficult variable to measure [see Friedman, Reed and Carterette, 1971]. In the present test sets, there was an attempt to make the largest (i.e., the most salient) changes at the end, rather than at the beginning, of the set, so that any tendency towards ceiling effects could be minimized. In Experiment 3, changes made in distractor sets for each prototype were all positional changes, which were large enough to be easy to discriminate, in order to achieve more control over salience of the details changed.

Conclusion

The present study indicated that motivation to remember does not help subjects to encode more information about stimuli in a recognition experiment. The study was not conclusive, however, since there appeared to be biases in confidence ratings in different groups, and since the scoring of memory accuracy was not completely satisfactory. The orienting task used in the present experiment, counting repetitions of types of items in the pictures, may have been facilitating in increasing accuracy scores, because subjects were able to perform as well with the orienting task as subjects given no orienting instructions and allowed to encode the pictures with freely chosen memory strategies. This could mean either that:

(1) subjects encode the same amount about a picture regardless of the attentional strategy they are using, or

(2) under certain conditions, subjects can encode more about a picture due to more efficient attentional strategies than subjects tested under different conditions.

The experiment in the next chapter explores these two possibilities.

CHAPTER FIVE

A COMPARISON OF ORIENTING TASK INSTRUCTIONS WITH STANDARD MEMORY INSTRUCTIONS IN RECOGNITION OF PICTURES

5.1 INTRODUCTION

There were many problems with Experiment 2. Experiment 3 attempted to provide a better test of the differential encoding hypothesis, so that the effect of instructions on memory for pictures could be better understood. At the same time, it was also intended to investigate more fully the exclusion set paradigm.

Once again, memory under intentional and incidental instructions was compared. In the present experiment, however, subjects were given a new orienting task, that of counting the number of "circles" in the pictures. This seemed to rely less on a semantic interpretation of the pictures than the task used in the previous experiment, where subjects had to identify classes of represented items in order to count repetitions. The circular content of the pictures could be varied, to determine the subjects' activities during the orienting task. The number of circles counted could be used as a crude measure of the level of attention given to the pictures, more circles counted signifying more encoding of the picture. Also, the circle-counting task seemed closer to the typefont-identification orienting tasks of verbal studies. The design of Experiment 3 was similar to the design of Experiment 2. One group of subjects was given only the orienting task with incidental memory instructions and another group was given intentional instruction with the orienting task. Instructions to this latter group emphasized the importance of the orienting task as a memory strategy. This was done to reduce the conflict between old strategies and experimental instructions which seemed to interfere with each other in the previous experiment. The orienting task used in Experiment 2 may have been facilitating for memory of the pictures. Because the circle-counting task seems easier for subjects to use, it should not have required processing at the same semantic level and therefore might have been less facilitative for memory than the task used in the previous experiment. It was hypothesized that, in this case, intentional instructions would lead to better recognition than incidental instructions, if awareness of the memory task is relevant for adequate encoding of stimuli.

An intentional group given no orienting task was also included in the experiment as a comparison "standard" memory group. The previous experiment had found no advantage of free coding when compared to an orienting task, but scores appeared to be lower in that experiment than in Experiment 1, in which a group allowed free coding of the same pictures was also given practice with the task before presentation of the stimuli. It was hypothesized that the superiority of the intentional group given immediate recognition testing in Experiment 1 over the intentional group in Experiment 2 was due to anticipatory encoding made possible by the practice trials given in the former experiment. Consequently, the intentional subjects in the present experiment were given free coding plus a practice trial to see if their memories would still demonstrate no advantage over memories of subjects in groups given orienting tasks. The hypothesis was that subjects who could anticipate the difficulty of the recognition task, as a result of practicing on one trial, would perform much better than subjects who were unaware of the amount of detail which would have to be remembered to achieve a perfect score. This is not an obvious deduction from theories which hypothesize that recognition is a simple activity. If encoding of pictures is some sort of automatic process, then attention to detail is as great with incidental or orienting instructions as with a practice trial. The only confounding factor might be subjects' "cheating", i.e., guessing how the sets are put together. Intentional subjects have one more trial on which to base their guesses about the sets. This will be discussed later as an error variable.

The intentional group also served to provide further information about the exclusion set recognition method as an alternative to standard memory paradigms. In this experiment, the problems of interference between members of a test list, standardization of the test procedure and response biases were dealt with more adequately, using improved procedures and a more closely controlled set of pictorial stimuli.

The stimulus set included one schematic face to see whether this picture would be remembered as poorly as other kinds of pictures. There is a great deal of literature on memory for faces, which has developed in parallel with studies in memory for other kinds of pictures. Few studies employing either photographs or schematized drawings include faces as well as other pictorial stimuli in the same list. Memory for views of the human face has received special attention because of the suggestion that memory for faces is specific to the right hemisphere. Studies with brain-damaged patients suggest that damage to the right hemisphere impairs recognition of faces much more than damage to the left hemisphere [Benton and Van Allen, 1970; Milner, 1968; Tzavaras, Hecaen and LeBras, 1971; Warrington and James, 1967]. There is some support for the notion that the deficits in recognition of faces suffered by patients with right hemispheric damage are a specific amnesia for one type of information, and this condition has been termed "prosopagnosia" [see Warrington and James, 1967]. Other evidence from experiments with normal subjects suggests that faces are a special class of visual stimuli, and hence may be remembered differently to other stimuli. Many researchers have shown that memory for faces has a high capacity [e.g., Galper and Hochberg, 1971]. However, memory capacity for faces is significantly poorer when faces are presented in the photographic negative rather than the positive [Galper, 1970; Galper and Hochberg, 1971] or when faces are inverted [Goldstein, 1965; Hochberg and Galper, 1967]. Faces are more impaired by inversion in recognition than are pictures of houses, airplanes, pictures of whole people [Yin, 1969], faces of dogs or buildings [Yarmey, 1971]. Thus it appears that faces, unlike ordinary pictures, are a special category of stimuli which require upright, positive presentation to be consistently remembered. A schematic face is tested in the present experiment, with recognition distractors based on an exclusion set scale of physical similarity, in order to compare faces with other types of pictures when testing is carried out under comparable conditions.

5.2 EXPERIMENT 3: EFFECT OF INSTRUCTIONS TO REMEMBER ON RECOGNITION OF PICTURES

5.2.1 Method

Subjects

Forty-five first-year Psychology students volunteered for the experiment, and were paid $2.00 for the 45-minute session. Fifteen subjects were assigned randomly to each experimental group, and both males and females were included in each group. Data from an additional four subjects was discarded because of an early technical failure.

Materials

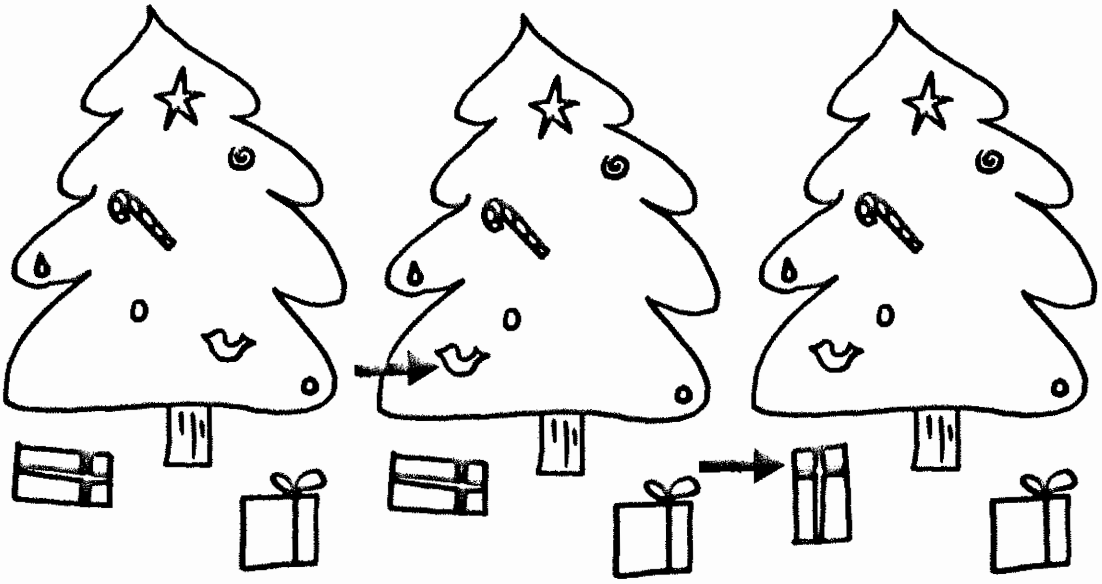
A set of 12 new stimuli was developed for this experiment. These were much simpler line drawings than those used in the earlier experiments. Some were based on the more complex drawings used in Experiments 1 and 2, and others were derived from children's books. The picture that was least satisfactory in drawing quality and detectable changes, as determined in pilot testing, was used as the practice item. For each picture, the test set was drawn using the same method outlined in Chapter Three. (There was one exception: the fifth variation of the face picture included *two* changes. This set is presented in Appendix 3.) All changes were cumulative so that each variation made was objectively more different to the prototype than its preceding variation. Only five variations were made for each picture in this experiment, making six items in each test set. It was difficult to include any more changes of equal difficulty in the pictures, since, in order to provide greater control over the type of changes made, the changes were all displacements of items in the pictures. Each picture was drawn on transparent plastic. Then identifiable pieces were cut out of the plastic so that each could be moved independently of all the others. The plastic pieces were arranged into the portrayed scene and this prototype was photographed. The first variation was made by moving a piece in the scene to create a noticeable change. This variation was photographed, and then successive variations were made until the fifth variation was completed. An attempt was made to keep displacement distances equal, but this was impossible in making coherent scenes for all pictures. Rather, changes made were simply large enough to be easily perceptible to observers. An example of one of the test sets is shown in Figure 7 (Picture 3). The remaining pictures are included in Appendix 3.

There was no attempt to define the number of circles present in each prototype. Indeed, the orienting task of counting circles in the pictures was developed after the stimuli had been chosen, because the pictures seemed to be more varied in their content of circles than in their content of any other geometric shapes. The purpose of choosing a class of shape to be attended to in each pictorial scene was to encourage subjects to see the picture as an array of lines without requiring them to process each picture semantically. However, forcing subjects to attend only to easily identified shapes (e.g., circles or squares) in each picture seemed to be too easy an orienting task. This would encourage intentional subjects to count the number or circles quickly and to use the rest of the presentation time to encode other kinds of information in each picture. By asking subjects to look for circles in each picture, without really defining the term "circle", subjects could be encouraged to attend to the picture during its entire presentation to examine each detail to determine whether it could be classified as a "circle". Thus the task would encourage both incidental and intentional subjects to attend to the picture in the same way. (In most studies employing a control group which is given the same orienting task as an incidental group and is told that memory for the stimuli is to be tested, it is assumed that control subjects are superior in motivation to incidental subjects, rather than merely superior in attending to more information in each stimulus.)

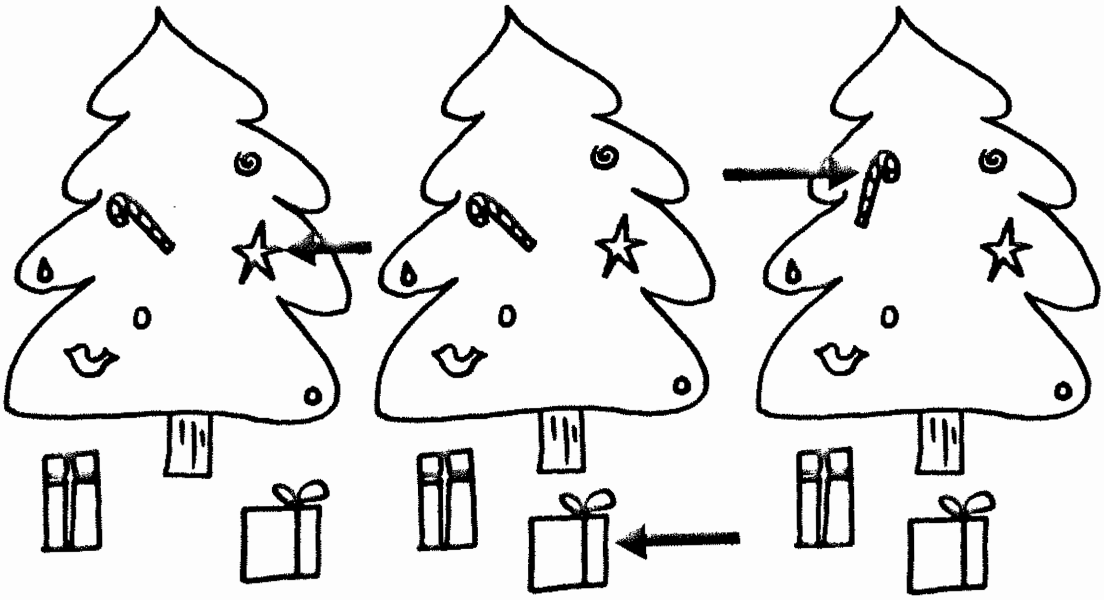
The pictures were all reduced to 35 mm black and white slides.

Apparatus

The prototypes were presented by a Kodak Carousel slide projector, Model S-AV2000, modified with a timing device to control presentation time and interstimulus interval. There was an internal shutter over the lens so that no light was projected during the interstimulus intervals. The test sets were projected by a Kodak Carousel slide projector, Model S-RA, positioned next to the other slide



A. Prototype **B. V1** **C. V2**



D. V3 **E. V4** **F. V5**

Figure 7. The test set for Picture 3 in Experiment 3 showing: (A) the Prototype, (B) Variation 1, (C) Variation 2, (D) Variation 3, (E) Variation 4 and (F) Variation 5. Arrows point to the detail changed in each variation.

projector so that both projected images onto the same area of a back projection screen, 48 cm by 32 cm. Pictures presented subtended a visual angle of approximately 12° by 18°. Subjects were seated approximately 150 cm from the centre of the back projection screen. Filters were positioned over the lenses of the projectors to reduce brightness. The room was dimly illuminated.

Procedure

Different instructions before the recognition task were given to subjects in the three experimental groups.

(1) Incidental: The experimenter gave the following instructions to subjects:

> "I am interested in the way we learn to look at a picture and see it as a 'scene' with meaningful objects as well as see it as a series of lines and geometric shapes. I have some pictures I want you to look at and I have something in particular I want you to concentrate on. I want you to try to find as many circles as you can in each picture. Now I'm going to show you a picture for a few seconds and I want you to count as many circles in it as you can while the picture is on. They don't have to be perfect circles; the drawings are done by hand."

The practice item was presented to subjects for five seconds. Then the subject was asked to report the number of circles found in the picture.

(2) Control: Subjects in this group were told initially:

> "This is a memory experiment. I have some pictures I'd like to show you and I want you to try to remember them. Afterwards I want to see how well you can recognize them by picking the originals out from a lot of similar ones. From studies I've done so far I've found that people have difficulty concentrating on a picture. I want you to use a special strategy I've devised for helping you to concentrate on each picture. I want you to try to find as many circles as you can in each picture."

After these instructions, subjects were given the same procedure for learning the circle-counting orienting task as subjects in the Incidental group, including the presentation of the practice item. Subjects in these first two groups did not have experience with the memory task while being acquainted with the orienting task.

(3) Intentional: Subjects in this group, like the Control subjects, were told that the task was a memory experiment, but they were not given an orienting task. Instead, they were given the practice item for five seconds and then shown the six items in its test set in a random order as a practice recognition trial. Subjects were told that only one picture in the set would be the one just seen and the others would all be slightly different from each other. For each picture shown, subjects made an oral "yes" or "no" response to indicate whether or not they thought each picture was correct (i.e., remembered). Subjects were not given experience with the confidence rating procedure.

After the initial instructions, Intentional subjects were told to get ready to remember "about a dozen" pictures which were to be presented sequentially for five seconds each. Incidental and Control subjects were told to prepare for the circle-counting task. The 11 test pictures were presented in an individually randomized order, with five seconds presentation time and 10 seconds interstimulus interval. The longer interstimulus interval used in the present experiment was chosen to allow subjects enough time to make a response on the orienting task. The memory stimuli were easier to remember than those in the first experiments reported, so presentation time was shortened to increase task difficulty. Intentional subjects remained silent during the presentation, while subjects given the orienting task said aloud the number of circles counted in each picture during the interstimulus intervals.

After presentation of the stimuli, all subjects were given the memory test. The task was explained to subjects who had been given the orienting task (and subjects who had been given intentional instructions only) by the experimenter in the following way:

> "Now I want to show you a whole series of pictures, all very similar to the ones you just saw. There will (again) be six pictures in each set, all slightly different, but only one will be the original one you saw. I want you to look at each picture and tell me whether you think it is one of the original ones you saw or not."

Subjects were asked to make a rating for each picture using the following scale:

(1) I am positive it is the original picture.

(2) I am fairly sure it is the original picture.

(3) I think it could be the original.

(4) I don't think it could be the original.

(5) I am fairly sure it is not the original.

(6) I am positive it is not the original.

Subjects given the orienting task were told,

> "Forget about the circles now and see if you can remember those pictures."

Subjects in the Incidental group were also told,

> "I wasn't trying to trick you by not telling you about the memory task. I wanted you to devote your whole attention to the circle counting. Now I want to get an idea of what you noticed in each picture."

Each stimulus set for the 11 test pictures was presented to subjects in a random order. Within each set the six test pictures were also presented randomly, one at a time. Subjects were given as long as they needed to decide on a rating for each picture. The rating scale was printed on a card in front of subjects and they said a rating aloud for each of the 66 pictures. When subjects rejected a picture (i.e., gave it a rating of "4", "5" or "6"), they were asked to give a reason for rejecting it, which was noted. Subjects were informed that their ratings did not have to be consistent and that they could change their minds during a set.

After recognition testing, subjects filled in a written questionnaire on their expectations about the experiment. This was the same as that used in the previous experiment. Two subjects in the Incidental group had anticipated the memory experiment and were rejected from data analysis.

5.2.2 Results

Effects of Instructions on Recognition of Pictures

The accuracy of recognition in the experiment can be scored by examining correct rejections or by measuring the hit rate. There is as yet no evidence to suggest that memory processes concerned with acceptance of an old stimulus are the same as those concerned with rejection of a new stimulus. So, for the moment, these will be dealt with independently.

Correct Rejections of Distractors:
Exclusion Set Scores

An exclusion set score for each subject on each picture was determined by noting the number of the lowest variation in the test set which the subject could reject with a correct reason. The exclusion set score was measured as six minus the number of the lowest variation in the test set which the subject could reject, so that scores ranged up to a perfect score of five when subjects rejected the first variation from the prototype. Subjects unable to reject any distractors were given a score of zero. Crucial to this measure is the subject's identification of an incorrect detail in a distractor, and such rejections were counted whether confidence ratings of "4", "5" or "6" were used. This was done so as not to disadvantage subjects cautious in using extremes of the rating scale. Also, because subjects saw each picture only once and may have failed to attend to a crucial detail on any one occasion, subjects were credited with rejecting a particular detail in all appropriate distractors even if they were not consistent in rejecting all pictures containing that detail. For example, a subject might reject only one picture in a set and give as the reason the detail changed from the prototype to the first variation. This subject would be credited with an exclusion set of five items no matter what variation the subject had actually rejected. This is a necessary procedure following the assumptions on which the exclusion set method rests. Because small changes are involved, accuracy results could easily be biased by subjects' failing to attend to details or their adopting a cautious response strategy if this correction were not used. Sometimes subjects rejected a picture with no reason or an incorrect reason, despite the probing procedure used in testing. These were difficult to interpret in terms of memory accuracy and were not counted as correct rejections.

Means and standard deviations of exclusion set scores for each instructional group are shown in Table 12. Results for each picture are shown in Appendix 3. A two-way analysis of variance treating the repeated measures on Pictures as a random factor was carried out on these scores and is summarized in Table 13. The main effect of Instructions was significant at the .005 level, while the main effect of Pictures was significant at the .001 level. The interaction of these factors was not significant. Newman-Keuls comparisons carried out on group means after the analysis (using the procedure outlined in Winer [1971]) revealed that the Intentional group performed significantly better than either the Incidental group ($p < .05$) or the Control group ($p < .01$), while the latter two groups did not differ significantly from each other. The comparisons for each picture showed that Pictures 7 and 9 were better remembered than many others, while Picture 8 was less well remembered than these and four others.[1] The other eight pictures were not significantly different in mean recognition scores. Because of the heterogeneity of the pictures, a t-test was carried out to determine whether the Incidental and Control groups differed in exclusion set scores on the eight consistent pictures. The Incidental group was still superior to the Control group, but the difference was not significant ($t = 1.3445$, $df = 28$, $p > .05$ for a two-tailed test).

Table 12

Means and standard deviations of exclusion set scores
for each instructional group: Experiment 3.

	Incidental	Control	Intentional
$\bar{X}$	2.4970	2.2545	3.2970
S.D.	0.6608	0.7723	0.9369

---

[1] Taking a 5% cutoff level of significance, Picture 9 was better remembered than Pictures 1, 2, 3, 6, 8, 10 or 11. Picture 7 was better remembered than Pictures 2, 8, 10 or 11. Picture 8 was remembered significantly poorer than Pictures 1, 3, 4 and 5, as well as Pictures 7 and 9.

Table 13

Summary of the analysis of variance
of exclusion set scores: Experiment 3.

Source of Variation	SS	df	MS	F	p
Between subjects	392.4767	44			
Instructions	98.1979	2	44.0990	7.0075	< .005
Subjects within groups	294.2788	42	7.0066		
Within subjects	1232.7273	450			
Pictures	169.6484	10	16.9648	7.1473	< .001
Instructions × Pictures	66.1577	20	3.3079	1.3936	NS
Pictures × subjects within groups	996.9212	420	2.3736		

To compare this exclusion set score with a more traditional analysis, an analysis of all rejections of distractors was carried out. Subjects were given a correct rejection score on each picture by counting the number of distractors given a "4", "5" or "6" rating of rejection and dividing this by the total number of "4", "5" and "6" ratings. This represented the proportion of distractors rejected, given the subject's tendency to respond negatively. Means and standard deviations of these scores for each Picture and each Instructional group are presented in Appendix 3, along with a summary of a two-way analysis of variance carried out on these scores. Replicating results of the exclusion set analysis, there were significant main effects of Instructions and Pictures, but no significant interaction. *Post hoc* comparisons of these scores showed that the Intentional group performed significantly better than either the Incidental or Control groups, and that these latter two groups did not differ significantly from each other. (Actually, two-tailed t-tests were carried out on the independent means of pairs of instructional groups, since no standard *post hoc* comparisons procedures yielded any significant differences between the instructional groups.) Newman-Keuls comparisons carried out on the means of scores for each picture indicated that only one picture was significantly different to others; Picture 2 was remembered more poorly than Pictures 4, 5 and 9 at the 5% level of significance. This does not correspond with the findings of the exclusion set analysis, although Picture 9 was better remembered than Picture 2 in both studies. Thus, though general trends in the two analyses were the same, the variance in scores contributed by the Pictures factor was produced by different pictures in the two analyses.

Hit Rate Analysis

Accuracy was also scored by examining responses to each prototype. A hit was scored by counting the ratings of "1", "2" or "3" given to prototype items and then dividing by the total number of these responses given in the set as a positive response-bias correction. Ideally, the accurate subjects would give a "1" rating to the prototype and ratings of "6" to all the distractors, and receive a hit score of 1.0, while subjects who reject the prototype would receive a score of zero. Means and standard deviations of hit rate scores for each Instructional group are shown in Table 14. Means and standard deviations of scores for each picture are shown in Appendix 3. A two-way analysis of variance with fixed effects factor of Instructions and the repeated measures on Pictures as a random factor was carried out on the data. The summary table (Table 15) shows that there was a significant effect of Instructions but no effect of Pictures and no interaction effect. Newman-Keuls *post hoc* comparisons carried out on the means scores revealed that the Intentional group was superior to the other groups ($p < .05$), but that the groups given an orienting task did not differ significantly from each other.

Table 14

Means and standard deviations of hit rate scores for each instructional group: Experiment 3.

	Incidental	Control	Intentional
$\bar{X}$	0.2278	0.2382	0.3986
S.D.	0.0825	0.0848	0.1350

Table 15

Summary of the analysis of variance of hit rate scores: Experiment 3.

Source of Variation	SS	df	MS	F	p
Between subjects	7.6469	44			
Instructions	2.9765	2	1.4883	13.3840	< .001
Subjects within groups	4.6704	42	0.1112		
Within subjects	45.7933	450			
Pictures	1.8124	10	0.1812	1.8322	NS
Instructions × Pictures	2.4467	30	0.1223	1.2366	NS
Pictures × Subjects within groups	41.5342	420	0.0989		

Conclusions about Memory Accuracy

All the accuracy analyses support the hypothesis that free intentional instructions with a practice trial lead to better recognition than orienting task instructions, and also demonstrate that knowledge of the memory task does not help subjects who are performing an orienting task while stimulus pictures are presented. Intentional and Incidental memory groups given orienting tasks did not differ in any analysis.

The finding of superiority of the Intentional group over the two other groups is ambiguous because the orienting task groups were at a disadvantage at the start of recognition testing. Intentional subjects received a practice trial so that they would be cued for noting appropriately small details in the test pictures, but at the same time they learned that the six pictures in the practice set were extremely similar and that cautious responding and careful discrimination between successive items was necessary. Orienting task subjects did not know this until the end of the first test set. To determine whether the strategy adopted during the testing procedure was the only reason that the Intentional group was superior, the data for each subject's exclusion set and hit rate scores were reanalysed excluding the first recognition test set. Scores on the remaining 10 sets were averaged and the three instructional groups were again compared, this time in a one-way analysis of variance. There was still a significant effect of Instructions for both the exclusion set scores ($F = 7.124$, $df = 2,42$, $p < .005$) and hit rate scores ($F = 15.023$, $df = 2,42$, $p < .001$), showing that the disadvantage of the orienting groups on the first test relative to the Intentional group was not the cause of their inferior memory scores. This result supports the hypothesis that intentional subjects given practice before testing encode more about the stimuli than subjects given no practice and instead given an orienting task.

To determine whether the various accuracy analyses were consistent, correlations across subjects between exclusion set scores and hit rate scores were computed. The rejection analysis was not included in this, as it was not independent of the exclusion set analysis, and was thought to contain the most inaccurate estimates of subjects' abilities. The mean exclusion set score was paired with the hit rate for each of the 15 subjects in each instructional group to calculate a Spearman rank correlation coefficient. The r_s for the Intentional group was 0.88, which was significant at the 1% level. The r_s correlations between hit rates and exclusion set scores were 0.47 and 0.46 for the Incidental and Control groups, respectively. Both correlations were significant at the 5% level. Since none of the correlations were high enough to suggest concordance of the two accuracy measures, the scores were not combined to derive a single accuracy score for each subject.

Orienting Task Responses

Subjects showed a variety of approaches to the orienting task. Subjects in the Incidental group counted 4.224 (S.D. = 1.6854) circles per picture, on average, while Control subjects counted only 3.000 (S.D. = 0.9706) on average. A t-test was carried out on the independent mean frequencies of circle counting for subjects in the two groups, pooled over all pictures. Subjects in the Incidental group counted significantly more circles than subjects in the Control group ($t = 2.437$, $df = 28$, $p < .05$ for a two-tailed test).

The face picture (Picture 7) was labelled by all subjects as having three circles, while the plant picture (Picture 10) had a range of counted circles from zero to 20. To see whether subjects who counted more circles remembered more of a picture shown, a Spearman rank correlation coefficient was carried out between circles counted and corresponding exclusion set scores for each subject in the two groups for Picture 10. This picture was chosen because of its wide response range on both scores. The r_s for the Incidental group was 0.46, which is significant at the 5% level, but the correlation of 0.43 found for the Control group fell short of significance at the 5% level.

Questionnaires

Results of the post-recognition questionnaire given to all subjects are shown in Table 16. None of the Incidental subjects reported trying to remember the pictures when they saw the original stimuli.

Table 16

Frequencies of responses to the post-recognition questionnaire: Experiment 3.

Question	Group		
	Incidental	Control	Intentional
1. Did you try to remember the pictures when you saw them?			
Yes	0	6	1
No	15	9	14
2. Were you surprised by the memory task?			
Yes	8	3	0
No	7	12	15
3. Rehearsal			
Yes	5	7	13
No	10	8	2

(Two subjects who did were rejected from the analysis and are not shown in the table.) In spite of this, many Incidental subjects also reported no surprise at the memory task when it was introduced. This seemed to be characterized by one subject's defensive, "I expect anything from psychologists", indicating that it was not inconsistent with answers to the first question. The Intentional subjects reported that they were all aware of the memory task, although one did not report making any effort to remember the stimuli on presentation, as they were "so easy". Some kind of active rehearsal was reported by 87% of these subjects. On the other hand, Control subjects reported that they found it difficult to use any remembering strategy while performing the orienting task, and three subjects more or less ignored the memory part of the instructions in order to concentrate on the circle-counting task. Control subjects appeared to rehearse the pictures in the intervals between pictures somewhat more than Incidental subjects seen during the intervals in testing. Incidental subjects often reported that they "went over their last responses in their minds" and thus actively rehearsed some information which the picture conveyed during the intervals between pictures.

The Face Picture

Previous studies have shown that female subjects are superior to males in memory for female faces [Cross, Cross and Daly, 1971; Goldstein and Chance, 1970] and that males are superior to females in memory for male faces [Laughery, Alexander and Lane, 1971]. No one seems to have investigated sex differences in memory for neuter, schematic faces. It was decided to test sex differences in this experiment despite the small number of male subjects available. No sex differences were found in the exclusion set experiment reported in Chapter Three which tested memory for pictures other than faces, but it was thought necessary to examine sex differences in recognition of the face picture, which may be remembered differently to other types of stimuli [see Scapinello and Yarmey, 1970]. In the present experiment, males' and females' exclusion set scores were compared for the face picture. Subjects who had been rejected for failing to follow instructions or because of equipment failure during pictures other than the face picture were added to increase the numbers of subjects tested. The Incidental and Control groups were combined to increase the number of males tested, after determining that these groups did not differ in recognition for either sex. The 12 males had an average exclusion set score of 2.75, while the 23 females had an average score of 3.17. A t-test based on the independent samples of males and females found that this was not a significant difference ($t = 0.9343$, $df = 33$, $p > .05$ for a two-tailed test). Males performed better than females in the Intentional group. The six males had an average exclusion set score of 4.67, while the 10 females had an average of 3.50. Once again, a t-test did not find that these means differed significantly from each other ($t = 1.6632$, $df = 14$, $p > .05$ for a two-tailed test). There were no sex differences found for memory of the face picture.

Exclusion Set Analysis of Confidence Ratings

Because complete data were available from all subjects, an opportunity was provided for a close examination of the exclusion set method. If the pictures in each test set provide a scale of similarity to the prototype, there should be a trend in the confidence ratings such that most dissimilar items are most often rejected. Figure 8 shows average confidence ratings given for each item in the test set, over all positions and for only the first list position, for each experimental group. The means and standard deviations on which these graphs are based, and data for each individual picture, are included in Appendix 3. Figure 8 shows that the Intentional group had an increasing tendency to reject distractors less similar to the prototype together with a high level of acceptance of the prototype. The Control group showed the same positive trend, but there was less acceptance of the prototype, while the Incidental group gave lower ratings of confidence to the prototype than to its first distractor variation, so that there was no positive trend relating confidence to the similarity of distractors to their prototype. Data for individual pictures (Appendix 3) also show that the Intentional group was more affected by similarity of distractors to their prototype than either of the groups given an orienting task. This gives further support to the finding that the Intentional group performed more accurately than the other groups. It also suggests that each successive distractor variation was perceived by subjects as less similar to the prototype, giving empirical evidence that similarity was manipulated adequately in the recognition task.

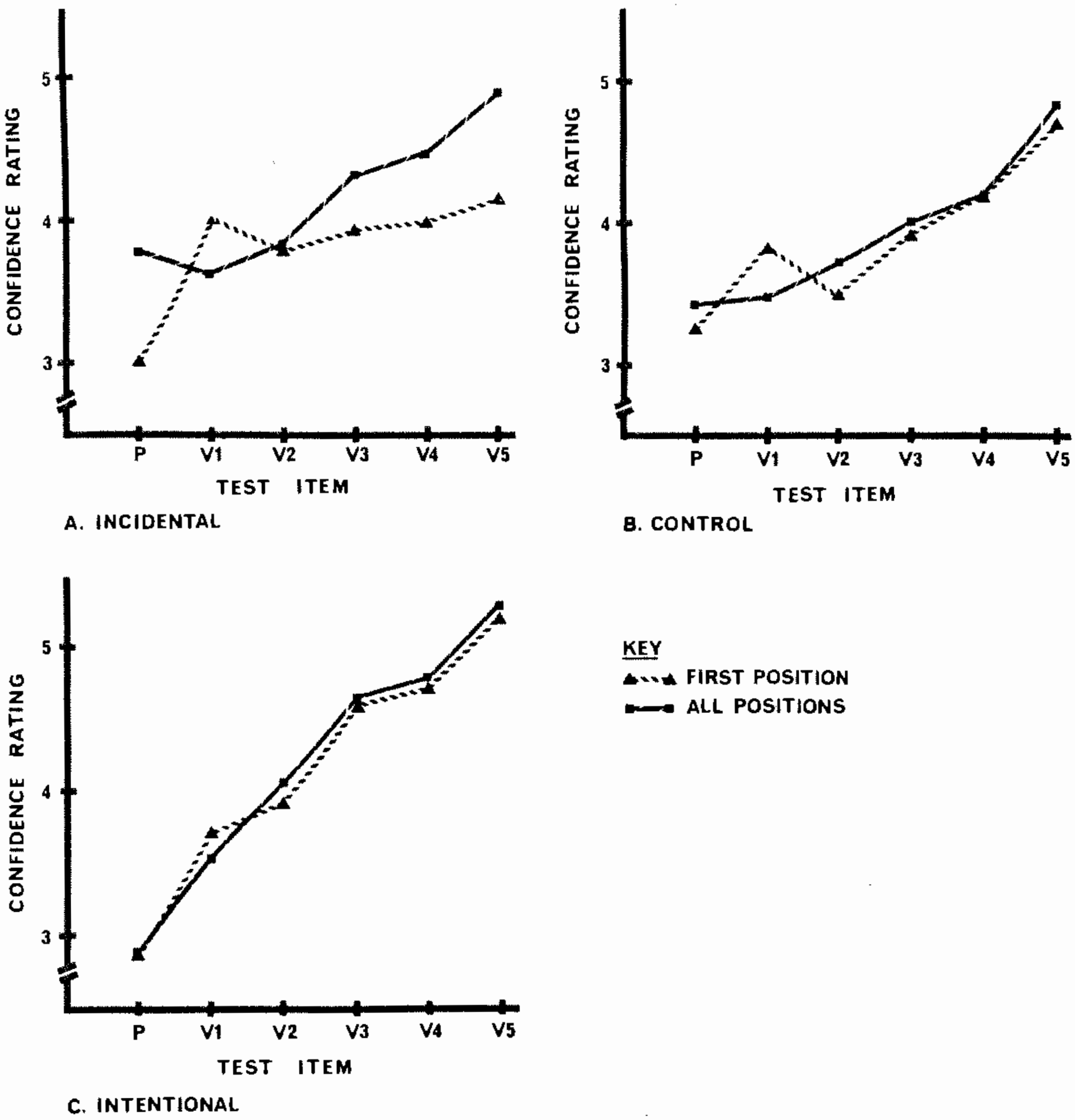
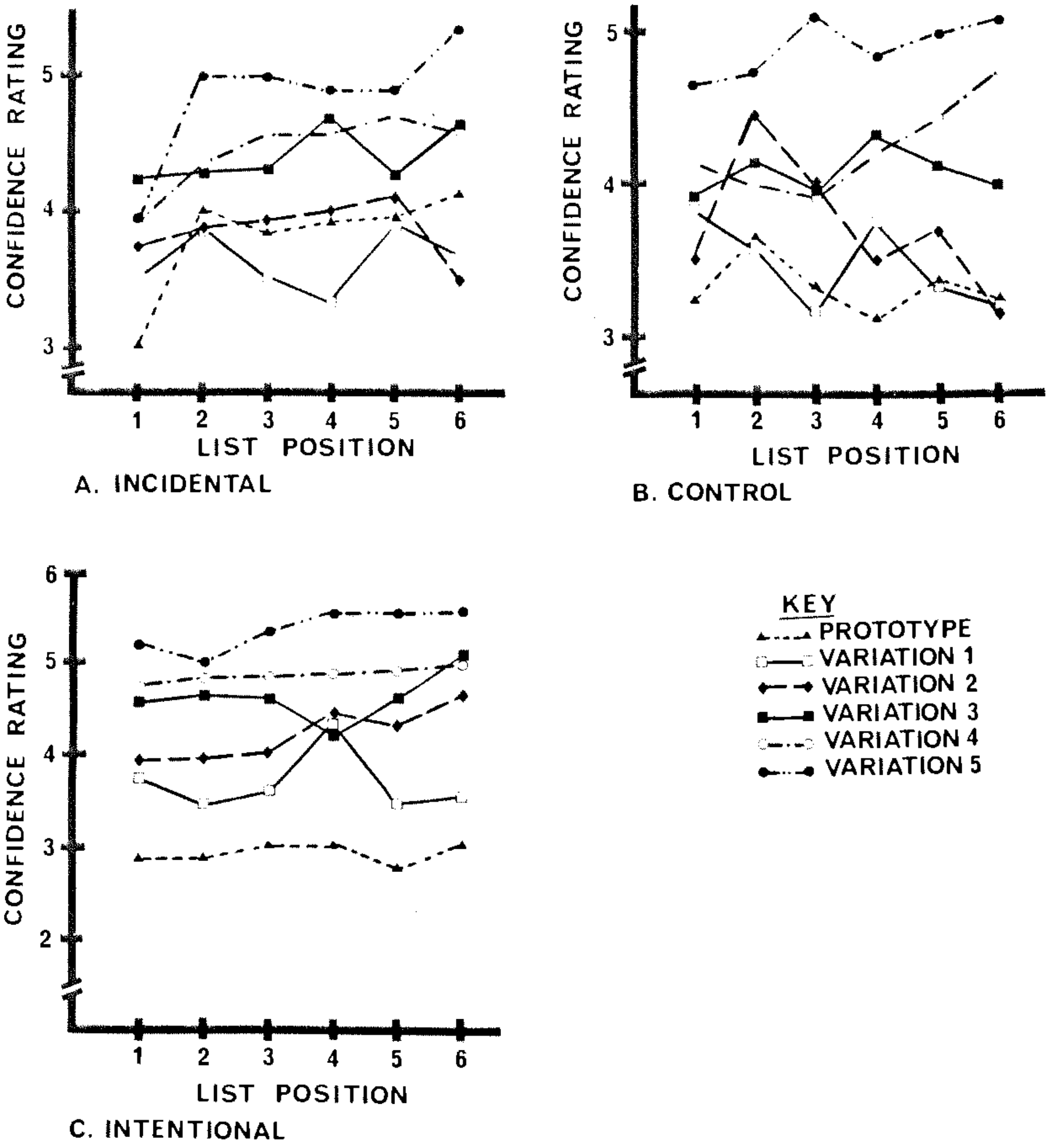


Figure 8. Confidence ratings for each item in the test sets, for all items regardless of their position in the list and for the first item in each list, for each group in Experiment 3.

One problem encountered in Experiment 2 was again found here. This was that subjects experienced interference between items in the stimulus sets and often complained that after seeing several very similar test pictures, they were totally confused about their memory of the original. To see whether confidence was a function of the position of an item in the set of six pictures presented to subjects, the confidence ratings were averaged over all pictures and all subjects in a group according to the position each item held in the set. Figure 9 shows confidence as a function of the position of each test item in the list for each instructional group. Means, standard deviations and frequencies of cells are shown in Appendix 3 for confidence ratings in each position of the list for each test item in the three instructional groups. The Intentional group showed no effect of interference due to the position in the list which each test item occupied. Confidence across list positions varied little for each item in the test set, and there was an increase in level of confidence with higher variations in the test set. The Incidental group also showed no consistent effect of list position on confidence ratings for a particular item, but there was a slight trend for the first items shown to be given more positive confidence ratings than others. This may be a cautious rejection strategy. Also, there was not the clear separation of confidence functions by list position for each stimulus that was found for the Intentional group. The Control group had the most bias in confidence on the basis of list position, since the prototype and its first four variations had ratings which differed according to their positions in the list, with no consistent trends. In this group as well as in the Incidental group, the prototype was given a higher rating if it appeared first in the list rather than when it appeared later in the list.

Unfortunately, it was not possible to carry out significance tests on these confidence data, since there were not equal numbers of responses to each picture in each position for all subjects. Once again, the Intentional group was the only one which was accurate in responding, giving ratings of positive acceptance to the prototypes regardless of their position in the list, and, in general, giving higher ratings to pictures less similar to the prototypes, regardless of their list positions. Thus, for this group, there were no primacy or recency effects in responding to the prototypes. The other groups' confidence ratings



A. INCIDENTAL

B. CONTROL

C. INTENTIONAL

Figure 9. Confidence ratings for each item in the test sets for each position in the list, for each group in Experiment 3.

were affected by list position, suggesting that these subjects do not have as much information in memory as Intentional subjects and are affected by biases, such as towards positive responding to the first item in the list.

5.3 GENERAL DISCUSSION OF EXPERIMENTS 2 AND 3

5.3.1 Effects of Instructions on Memory for Pictures

The results of Experiment 3 extend and replicate those of Experiment 2. Again, no differences in recognition between intentional and incidental groups given an orienting task were found, even though a new orienting task, different stimulus pictures and different recognition scores were used. Thus, when subjects are given an orienting task, their knowledge of the memory task to follow is irrelevant. These subjects will remember as much when they are preparing for a memory test as when they are not. Experiment 2 showed that intentional instructions only did not lead to recognition performance superior to intentional instructions when an orienting task was given. Thus, subjects appeared to remember pictures equally well no matter what activities they carried out during encoding. However, in Experiment 3 an intentional group given a practice trial performed better than either an intentional or incidental group given an orienting task. This was true even for recognition scores based on the last 10 sets tested in each group, discounting the first test set. This means that the advantage gained from the practice trial was not due simply to responding strategies or experience with the task. In fact, these subjects continued to learn about the task on the first test trial, since they were not given experience with the confidence rating scale on the practice trial. Thus, recognition of pictures *is* affected by the encoding activities which the subject carries out during a picture's presentation, and instructions about the nature of the memory task can affect subsequent memory accuracy.

The Intentional subjects in Experiment 3 recognized the prototypes more often and rated them with more confidence, and rejected distractors more often, especially if they were less similar to the prototype, than subjects given an orienting task with or without intentional instructions. Clearly, these subjects had encoded more about the pictures than the subjects given an orienting task without a practice trial. This is evidence for differential encoding in recognition of pictures and refutes the hypothesis that pictures are processed automatically to the same degree regardless of the subject's encoding activities. Adequate knowledge of the amount of detail which must be remembered can lead to performance superior to that of uninformed subjects.

A possible criticism of the findings of Experiment 3 is that Intentional subjects could notice the small details changed in the practice test set and so be cued to attend to the details in the stimulus list. It can perhaps be suggested that Intentional subjects were cued only to notice positions of items in the pictures, rather than to notice more details about each item. However, with only one practice trial, it is unlikely that subjects noticed that only positions of items were changed. Several subjects in the Intentional group spontaneously asked during the practice trial, "Are you sure this is a different picture?" Subjects did not always notice detail changes from one picture to the next, though presumably they were better at this on their second trial. Further evidence that Intentional subjects were looking for all kinds of detailed information in the stimulus pictures comes from the fact that only one of these subjects said that he had noticed that "mostly" positions were changed in the pictures. On the other hand, subjects in every group made incorrect rejections of pictures on the basis of content rather than position information. Subjects' failure to notice how the distractors were varied was probably due to their expectation that content changes would occur and their knowledge that there were differences between test pictures which they had not noticed. It is most unlikely that subjects in the intentional group given a practice trial performed better than all other groups on the basis of their test-wise sophistication. These subjects were allowed to develop the most relevant and specialized encoding strategies needed to remember the pictures and this led to their superiority in memory accuracy. Pictures are not encoded automatically under all circumstances; some strategies lead to more efficient encoding and better memories than others.

Experiment 2 contained a group given intentional instructions without an orienting task or a practice trial. The motivation to remember in this group, unhindered by a conflicting orienting task, did not lead these intentional subjects to recognize more pictures than intentional subjects given an orienting task. The results of Experiments 2 and 3 taken together suggest that subjects who are attending to details in a picture, for example, by noticing repetitions of items or by counting circles, remember as much as subjects gazing freely at the pictures and trying to remember them. Thus, as long as subjects attend to the pictures, they may remember quite a lot about them even if they are not trying to remember them. However, they will not remember as much about a picture as subjects who are cued to noticing small details in the pictures, since they remember less than intentional subjects given a practice trial. The practice trial was actually a difficult discrimination task in which subjects had to differentiate between successive items. Presumably subjects cued by the discrimination practice were better able to encode information from the pictorial stimuli than uncued subjects. However, there is no evidence that motivation *per se* leads to better encoding strategies.

Since there were similar patterns of confidence ratings and memory accuracy between the Incidental and Control groups, it might be suggested that these groups did not perceive the pictures in different ways. In other words, the independent variable of instructions may have been unsatisfactory. Theoretically, the Control subjects' awareness of the memory task should have motivated them to attend more carefully to details in each picture, to process information from the pictures other than that which is required by the orienting task and to rehearse in the ISI's between pictures. This might have made the Control subjects remember more than the Incidental subjects concerned only with the orienting task. However, Control subjects may have ignored the instructions concerning the memory task. Evidence for this comes from Control subjects' answers to the post-recognition questionnaire: 30% of Control subjects in Experiment 2 and 20% in Experiment 3 reported that they did not "try to remember" the pictures during their presentation, presumably because of the demands of the orienting task. Other evidence is the report by Incidental subjects that they rehearsed in the ISI's nearly as much as Control subjects. It seems, in this case, that they were simply thinking retrospectively about their responses on the orienting task. These findings suggest that the Incidental and Control groups carried out the remembering instructions in a similar fashion.

However, there is one piece of evidence that demonstrates that Control subjects in Experiment 3 carried out the orienting task in a different way to Incidental subjects. Control subjects counted significantly fewer circles than Incidental subjects, on average. It appears that Control subjects spent less time on the orienting task, or used a more stringent criterion for accepting an item as a circle. It is obvious that they performed adequately on the orienting task, since two pictures were given unanimous circle counts by all subjects in the Control group. But presumably the Control subjects were aware that information other than circles had to be attended to in each picture for the memory task. Despite their different attentional strategy, they remembered no more about each picture than Incidental subjects. Thus, this strategy was not effective in increasing encoding. Motivation to remember pictures does not appear to lead to subjects' use of more efficient encoding activities than would be employed when subjects are not motivated to remember.

5.3.2 The Orienting Task: A Useful Independent Variable?

This study was intended to be a first step in a more detailed study of instructions to define a variety of orienting tasks to measure different levels of processing. Unfortunately, this would have relied very heavily on instructions as the independent variable. The single orienting task used in Experiment 3, the circle-counting task, was not the same for all subjects for all pictures. There was a wide range of circles counted for each picture in the orienting groups. This would be expected if pictures varied in their circular content, which was undoubtedly true, but subjects were consistent in their circle counting with some pictures, and not with others. The orienting task may have been ambiguous for these pictures.

Perhaps the ambiguity of the orienting task to many subjects was the reason that there was no clear relationship between the number of circles counted and recognition accuracy. There was a correlation of about 0.45 between recognition exclusion set scores and the number of circles in Picture 10, the picture with the largest standard deviation in circle counting. Obviously there would be a lower correlation in other pictures with little or no range in orienting scores. A more convincing demonstration that more counted circles did not lead to greater attention to the pictures during encoding is provided by the comparison of Pictures 6 and 7 in the Control group. All these subjects gave the face picture a circle count of three and the house picture a count of one. Despite the consistently different attention paid to the two pictures as indicated by the orienting task, none of the accuracy measures found differences between these two pictures for any instructional group, though means for each measure are in the right direction.

In a way, this single orienting task does provide a measure of processing at different degrees of elaboration. Perhaps subjects counting circles in a picture still have part of the presentation time of the picture to use for more general encoding, since the circle counting task may be a quick, efficient procedure. The two subjects who counted 20 circles in Picture 10 probably had no time in the five seconds of picture presentation to process anything except the requirements of the orienting task. Thus, it could be said that subjects counting one or three circles in the pictures still had the maximum time needed to use elaborated encoding as well as to perform the orienting task. The purpose of the orienting task in Experiment 3 was not to define different levels of processing, but its results strike some cautionary notes in the use of strategies to define deeper or more elaborated encoding. Firstly, there is the tentative indication that more processing at one level (i.e., more circle counting) does not always lead to better encoding for later recognition than less time spent in processing. This idea corresponds with a fashionable hypothesis in the verbal depth-of-processing literature [see Craik and Tulving, 1975] that it is the quality of encoding rather than the amount of time spent in encoding which is important.

A second and more fundamental problem in the area is to define the differences between processing strategies due to different instructions. In Experiment 3 the orienting task instructions were not made more explicit, because the instructional manipulation would be worthless if the task were too easy and subjects had too much time left over to encode the picture as they might in a normal memory task. Thus, some subjects interpreted "circles" as anything vaguely rounded, while others counted only closed-figure, unambiguous circles. The problem could be overcome by using unambiguous arrays of geometrical shapes. But the next problem would be to define another orienting task to indicate differential processing of this picture. "Physical" processing is easy to define: circle counting could be compared with rhombus counting, colour naming or corner classification. Defining any other level of processing encounters immediate difficulties, however. Bower and Karlin [1974] were criticized earlier for using judgments of sex, honesty and likableness of faces without attempting to define how subjects carried out these ratings, or what they meant in terms of processing. Yet any nonphysical ratings applied to pictures become bogged in circular assumptions about levels of processing. Thus a finding of a significant difference between two groups given different instructions, allegedly at different levels of semantic analysis, is used as support for differential processing. Processing may be different, but until the processes are defined and understood, these results will be virtually meaningless. Once again, pictorial studies point out problems not encountered in verbal studies. Linguists have distinguished between the deep structure of language as opposed to its surface structure for some time. But pictures are difficult to perceive without processing their semantic "message". It is doubtful whether the subjects in orienting groups in Experiment 3 would have had difficulty in describing global aspects of pictures seen. It seems implausible to suggest that processing was carried out on only structural aspects of pictures. Judgments about circular items were very involved for some subjects; semicircles were counted because they were "tyres" and even a square shirt cuff was labelled a circle because the subject was inferring circular properties based on a correspondence with a real world item. Thus, a decision about an item's circularity did not depend solely upon geometric contours in the drawings, but included semantic analysis of some kind.

5.3.3 The Exclusion Set Paradigm: Experiment 3

An attempt was made in Experiment 3 to create successive variations equal in amount of physical change. This was done by moving items in the array, rather than by changing the structure of items. It was hoped that this would produce changes resulting in equal amounts of global alteration to the picture. It was not feasible to show judges every possible combination of details which could be changed for each variation and ask them to make ratings for each possible picture in order to choose distractors varying from the prototypes in equally-spaced steps. This is because after the first variation has been chosen, the number of items available to produce the next variation is reduced by one. For the last variation, the degrees of freedom are so reduced that it may be impossible to control the change made to this distractor. This problem might be overcome by using a different type of stimulus. Arrays of unambiguous items, such as geometric shapes, could be used as prototypes and altered using only physical distance as the basis of change for each variation. This would restrict the task to memory for position and offer no evidence about memory in general. The present experiment was not really so restricted, since few subjects noticed that changes were made only on one dimension. Also, in these complex, realistic scenes, the positional changes made to distractors created globally different pictures. This may not be true for geometric arrays, which are not organized into scenes. The present stimulus sets seemed to be an important compromise between controlled stimuli and meaningful scenes.

The Intentional group in Experiment 3 provides a standard memory group on which to test some of the assumptions of the exclusion set method. Figures 8 and 9 showed that, in general, confident ratings of rejection became more common when distractors less similar to the prototype were shown, no matter what position the picture held in the list of six items. The finding that Incidental and Control subjects did not show clear trends between confidence and distractor similarity and were somewhat affected by list position does not invalidate the method. Such performance would be expected from subjects who do not remember much about the pictures; they would not be able to discriminate between correct and incorrect pictures and would be influenced by response biases. For example, their tendency to be positive about the first item in the list was probably due to cautiousness in responding because subjects did not notice critical details until they had a chance to observe differences between successive pictures. Alternatively, these subjects were not responding at all accurately and substituted a memory for the first picture shown for the decayed memory of the original. In contrast to this, confidence rating data from subjects in the Intentional group give support to the assumption that each variation in the test sets more changed from the prototype is more likely to be rejected in a recognition decision because it is less similar to the original. Intentional subjects' responses indicate that test pictures were perceived as items on an ordinal scale of similarity to the prototypes. This supports the validity of the construction of test sets using the exclusion set method.

Future work with the exclusion set method could lead to improvement in its design. At present there is little understanding of the way pictures are perceived. This makes it difficult to control perceptual similarity of pictures by physical means, which would provide the ideal objectivity. Also, in the future, different types of changes could be compared in producing test sets. Positional changes are easier to control than changes such as deletion, addition, substitution of a similar item or structural distortion, but it is not known whether these changes tap memory for different types of information or whether valid distractor sets could be made by sampling more widely from a variety of types of changes. The exclusion set method can be used to test memory for particular types of change as well as global memory for pictures through use of a variety of distractor changes. This would be useful in investigating the kinds of information most salient in pictures, and how they interact with each other.

5.3.4 Conclusion

The findings of Experiments 2 and 3 suggest that awareness of a memory task alone does not lead subjects to remember more about pictures than subjects unaware of the real nature of the task and who are encouraged to attend to the pictures by carrying out some irrelevant activity. This finding has implications for studies which examine effects of various orienting task activities on subsequent memory, since there is no need to include both intentional and incidental groups given the orienting tasks.

The finding that subjects given a practice trial before stimulus presentation remember more than subjects given no practice and instead given an orienting task during presentation provides evidence that encoding of pictures (including the organization of retrieval plans) is a complex, active process. Encoding may be more elaborate or relevant under some stimulus presentation conditions than others, which suggests that recognition of pictures is not an automatic, one-step process.

The exclusion set method was given some construct validity by confidence ratings of the standard memory group in Experiment 3 which showed that distractors less similar to the prototype were given more positive rejections in ratings. This also demonstrated that the differences between pictures were perceptible. The method could be improved in future studies by controlling the amount of change in distractors and by ensuring that changes are a random sample of possible changes. Types of changes could be compared and controlled, and the salience of particular changes could also be examined. Further development of the method itself could extend present knowledge of the processes involved in memory for pictures.

CHAPTER SIX

THE INFLUENCE OF TEMPORAL PARAMETERS ON MEMORY FOR PARTS OF PICTURES AND THEIR COMBINATIONS

6.1 THE FRAGMENTED MEMORY EFFECT

Experiments 1, 2 and 3 used the Exclusion Set approach to investigate memory failure due to loss of remembered detail. No other types of error, such as distortion of memory or transference of detail from one picture to another in memory, were investigated. Yet other types of error have been documented in studies of recall of pictures (Chapter Two). One of the most interesting errors in recall is the tendency for subjects to remember a single item in detail, but to remember it as having occurred in an inappropriate context. For example, in the preliminary study described in section 2.1.2, one subject interpolated a representation of a lamp into his reproduction of a picture representing a dining room scene [Figure 1(d)], when the only lamp presented to the subject had been in a picture representing a lounge room scene. Some early researchers also noticed such confusions in their subjects' reproductions from memory. Bartlett [1932] noticed "transference" of detail from one picture to another in subjects' reproductions of shapes as well as their verbal recall of male faces. Similarly, the most common error in subjects' drawings from memory noted by Gibson [1929] was that two of his stimulus shapes could be "blended" or confused into a single figure.

The occurrence of confusion errors in subjects' reproductions of pictures from memory has important implications for a theory of memory for pictures. It suggests that parts of a picture can be accurately remembered even though information about the whole picture is forgotten. This is surprising if pictures are encoded automatically as a Gestalt. The finding that instructions have an effect on recognition of pictures [Experiment 3, Chapter Five; also, Tversky, 1973, 1974] provides evidence that encoding processes vary with different conditions, and hence that they are not automatic and predictable. There is still an implication, however, from some studies, that pictures are remembered as wholes, memory being an all-or-none process. This implication is obvious in studies employing free recall (using words or drawings as the responses) as a measure of memory for pictures (sections 2.1.3 and 2.3). The possibility that half a picture could be remembered is not considered in these studies, but there is evidence that this is the case. Experiments employing the Exclusion Set method (Experiments 1, 2 and 3) suggest that subjects' memories of a picture can vary in accuracy from a perfect memory to memory for a few details. Given this as a characteristic of memory, the finding that subjects can remember fragments of pictures seen without remembering their context and make confusions between pictures is perhaps not surprising. This finding may be termed the "fragmented memory effect" since it appears to be an important aspect of memory. The effect has so far been documented only qualitatively in studies of recall of pictures. One purpose of the experiments to follow was to demonstrate the occurrence of this effect empirically in a study of recognition memory.

A preliminary study was carried out to determine whether confusions of items between pictures could be demonstrated in a recognition task.

6.1.1 A Preliminary Study of Confusions in Recognition of Pictures

A recognition task was devised in order to determine whether subjects can remember parts of a picture and yet be unable to remember the context in which these were presented. Distractors in the recognition set were created so that they were identical to the original stimulus with the exception of one detail, a particular item in each portrayed scene. In each distractor picture, a detail similar to one in the original picture was substituted for it. Two distractors were made for each stimulus tested, one containing a completely new detail in place of the old one, and one containing a similar detail from one of the other stimuli which was not tested. For example, in Figure 10 the picture at the top is the original item that was shown during presentation. The picture below is identical except that the lamp is changed. This lamp had actually been presented in another stimulus not tested in recognition (see Picture 3, Appendix 1). It was thought that subjects who made a false choice of this distractor in recognition might have remembered the lamp without remembering the context in which it was originally presented. Finally, at the bottom of Figure 10 is a picture identical to the original except for the substitution of a completely new lamp with which the subject has never been presented. By comparing the relative frequencies of subjects' false recognitions of the two distractor pictures, it is possible to determine whether pictures are rejected on the basis of any incorrect part or whether subjects make false alarms to pictures containing a remembered item in an inappropriate context. Subjects with accurate memories should be able to reject both distractors with equal frequency. However, if subjects can remember parts of pictures seen (e.g., the lamp in the bedroom scene) and fail to remember any contextual cues which aid them to determine which parts were presented with which in a particular picture, then distractors with a part substituted from another stimulus will be chosen more often than distractors with a completely new part. This would demonstrate the fragmented memory effect.

Method

Materials: Six complex black and white line drawings were used as the stimuli. Each picture represented a scene containing meaningful objects similar to the pictures used in Experiment 3 (Chapter Five). Only three stimuli were tested for recognition, while the other three pictures provided a source of details which could be substituted for a similar detail in one of the tested pictures to create a distractor with a "transferred part". Two distractors were made for each picture tested. In the transferred-part distractor, the taps in a drawing depicting a bathtub were replaced by taps which had appeared in a picture of a kitchen which was not tested. A second distractor was made by replacing the taps in the original with a completely new drawing of taps. To create the transferred-part distractor for a picture representing a

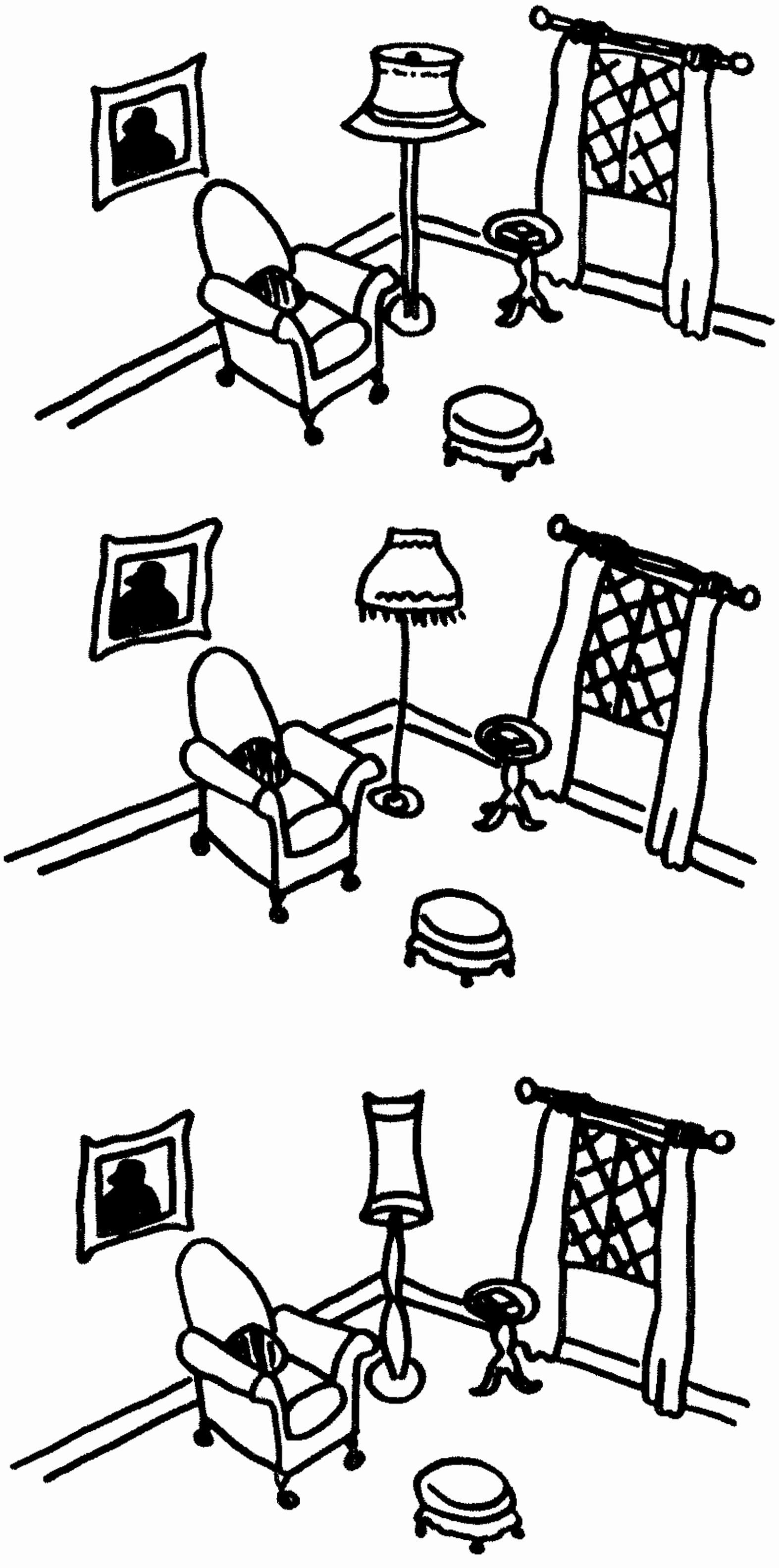


Figure 10. Three test pictures used in the preliminary study of confusions between pictures. The picture at the top is the prototype which was presented as a stimulus. The other two pictures are distractors containing a different item. The lamp in the central picture was presented in another stimulus picture depicting a bedroom (see Appendix 1), while the lamp in the bottom picture is a new item which does not appear in the stimulus set.

campfire, the pot portrayed in the original was replaced by a bowl which had appeared in a stimulus picture representing a dinner table: For the second distractor, the pot was replaced by a differently shaped pot. All pictures were hand drawn with a felt pen into a large booklet. The six stimuli were drawn on separate pages, enabling successive presentation. Each test picture was either preceded or followed with the non-tested picture which was the source of the item substituted in its recognition test set. It was hoped that this would maximize any tendency to confuse pictures. Order of presentation of pairs and position of the test picture in the pair were randomized. The three test pictures in each set were also drawn on successive pages, but the order of the members in each triad and the order of triads in the list were randomized. The test pictures were redrawn for the test list so that the subject could not choose on the basis of minute errors of draghtsmanship.

Subjects: Seventeen Psychology students and colleagues (six females) volunteered for the 10-minute session.

Procedure: Subjects were tested individually, seated at a desk opposite the experimenter. Subjects were told that they would be shown six pictures which they should try to remember. The booklet was placed in front of the subject and the experimenter turned a page over every 10 seconds, timed by a stopwatch, so that the six pictures in the stimulus list were presented in the same order for all subjects. After the stimulus pictures had been shown, subjects were told that they would be shown sets of three pictures, all slightly different from each other, and that they were to give a rating for each picture to determine whether it was the original or not. Subjects were given a six-point confidence rating scale printed on a card. This was the one that had been used in Experiment 3 (Chapter 5). The set of test pictures was then shown to subjects. Half the subjects were shown the test pictures in one order, while half received the reverse order. After all pictures were shown, if subjects had not noticed how the pictures in any set differed, they were again shown this set and the experimenter pointed out the crucial detail in each picture. On the second presentation, subjects were allowed to change their confidence ratings if they wished.

Results

The original pictures were recognized correctly on 75% of occasions and rejected falsely on 25% of occasions. Distractors containing an incorrect detail from another picture in the stimulus list were falsely recognized on 41% of occasions, while completely new distractors were falsely recognized on only 18% of occasions on which they were presented. Average confidence ratings were 2.9 for the original pictures, 4.1 for the transferred-part distractors and 4.8 for the distractors with a completely new part. Subjects had no difficulty discriminating correct from incorrect items, but distractors with details that had appeared in other pictures were more difficult to discriminate from originals than distractors with completely new details. In one or more of the recognition sets, 41% of subjects were unable to reject the original picture or the distractor with a part from another stimulus, while rejecting the distractor with the completely new part. Thus, these subjects were uncertain of the context in which they had seen particular details, but they did have information about these details, since they could reject a picture containing a completely new part.

Discussion

Although subjects demonstrated good memory for the pictures in this study, they were more confused by distractors containing items from another picture seen than by distractors containing never-seen items. This demonstrates that there was some breakdown in encoding of pictorial scenes, which led to confusions between pictures. Apparently, individual details in a picture can be remembered without any attendant global information about the picture in which they were seen or any contextual cues which could link the detail to another detail seen in the original. This suggests that the fragmented memory effect is an important memory phenomenon which could be investigated in experiments employing recognition methods.

The choice of stimuli in the preliminary study was unsatisfactory. The use of realistic scenes offers several problems for the study of confusions between pictures:

(1) it is difficult to generate sets of stimulus pictures with interchangeable items, i.e., contextual cues are likely to enable the subject's correct selection of the original;

(2) all pictures in a recognition set may not be drawn equally skilfully;

(3) items interchanged between pictures should not be able to be given distinctive labels which could act as retrieval cues when no visual information is remembered.

This final criticism is a difficult one to overcome when stimuli used are representational drawings, since some subjects will always be able to use verbal strategies with realistic scenes. Because of these considerations, difficult-to-label abstract shapes were chosen for the series of recognition experiments that follow, rather than drawings of scenes. This allowed simplification of the stimulus materials. The stimuli were composed of two parts: a symmetrical configuration made of six line segments enclosed within a bi-axially symmetrical 24-sided polygon. In free viewing, the inner form is characteristically seen as a figure lying on the outer form which acts as a background. The inside forms and outside forms could be combined into numerous varieties of two-part shapes, thus making it possible to generate large sets of stimuli with interchangeable parts. At the same time, the stimuli were difficult to label, and not too easy for subjects to remember even though they were composed of only two major parts. The stimuli could be controlled for their physical characteristics and were novel enough to prevent subjects' reliance on overlearned memory strategies.

6.1.2 Theoretical Implications

The finding that parts of different pictures can be confused is interesting in its implications for memory. Pictures are usually discussed as if they were unitary "events". According to a current view, an "event" is the unit of memory in real life, which is some experience ("episode") of the person which is stored in memory with an autobiographical reference [cf. Tulving, 1972]. If a picture is stored as a coherent event, it is surprising that its components can become disconnected and attached to components of other pictures in memory. The confusions found in the preliminary study suggest that the encoding of a picture as an event might be a sequential process such that, if there is not time to complete encoding, the parts of the picture may be registered in memory without their context. Clearly, the structure of a picture is not grasped instantaneously. The encoding process can be postulated as a series of successive attentions: to the global organization of the picture, to individual components, to details in the components and to relationships between them. Neisser [1967] and others have suggested that attention to an object may include an initial global impression, but that focussed attention is restricted to a subarea of the whole. If more than one focussed attention needs to be made to perceive and encode an event, then it is perhaps logical that these attentions are of a particular duration in time (not necessarily of fixed duration) and follow each other in a sequential order. Thus the experience of entering a room and glancing at different parts of it before registering an impression of the whole may be due to a more general process of successive attentions in constructing an event during perception. The necessity for this encoding sequence, which consists of activities of limited scope and duration which follow one another, lies in the assumption that the capacity of processing operations which can be undertaken at any time is limited. This idea that the amount of information which can be attended to at one time is finite has a large amount of supporting experimental evidence [see Moray, 1969].

It has been well documented that, in perceiving, subjects scan a picture using eye fixations to different parts of the scene. In a review of some of this literature, Kolers [1973, p.26] suggested,

> "Our subjective sense when we look at a picture ... is that we see the whole scene all at once; but the data indicate that we are examining those scenes only in small pieces, and over time."

Eye movements are no doubt important in the attention process, but it might be unwise to call each eye movement an act of attention which leads to the encoding of a certain amount of information. There is conflicting evidence about the importance of numerous eye movements in recognition of pictures [Loftus, 1972, versus Tversky, 1974] and there may not be any correspondence between patterns of eye fixation in perception and recognition testing [Locher and Nodine, 1974]. It may be more parsimonious at present to speak of successive attentions to a picture, which are not tied to a particular number of eye fixations. In the present context, an "act of attention" is a theoretical construct used to explain a process in event encoding.

Some testable hypotheses about memory encoding emerge from this interpretation of the confusion data. With brief stimulus exposures, there may be less opportunity for multiple acts of attention to an event. As a result, only a few parts of the stimulus will be remembered and there may not be an integrated memory of the whole. Before an event can be fully comprehended, all parts of the event must be scanned, so that all processes of attention necessary are completed. Thus comprehension of an event is a conclusion of some kind about what has been perceived. The perception of the event as a unity may not necessarily require a specific attentional act to terminate processing and register the event in memory. Rather, the perceiver may be constantly attending to different parts of the environment in successive acts of attention and consciously experiencing events as a sequence of fragmentary views. Because of this, many parts of a picture may be remembered which have not been consciously encoded into a context. This produces the fragmented memory effect. Sometimes, for example, with fast stimulus presentations, acts of attention to parts of stimuli from different pictures may not be punctuated by the perceiver's awareness of each individual event. Perceivers may encode only a succession of parts, not a list of well-articulated whole events. Thus, parts must be remembered in a context with other parts before an event can be comprehended. The context may be provided by some global notion extracted early in processing and "tagged" onto each part to which the perceiver is attending.

This view of memory complements well the view expressed by Lockhart *et al.* [in press] which hypothesizes that events are recorded as a continuous, chronological episodic trace without time tagging or other built-in reference systems. Memory is accessed by sorting through traces proximal in time to find contextual cues about events to be reconstructed. During perception, successive acts of attention may lead to individual memories about the parts attended to which can,be then constructed into events or left as fragments which are not remembered in the context of a specific event. Event encoding, that is, consciously concluding that a series of parts go together to make a particular experience (e.g., the experience of viewing a picture) may be improved when events are separated in time, so that parts attended to are encoded in a specific temporal context. For example, when pictures are presented sequentially to subjects with long presentation times and long interstimulus intervals, it should be easy for subjects to encode each picture as a separate event. But when pictures follow one another rapidly, it may be difficult to link the parts given focussed attention in each picture together as a whole. The separate acts of attention to different areas of pictures may be well remembered, but memory for each whole scene may be poor. It was hypothesized that fragmented memory would occur more often when shorter presentation times or interstimulus intervals are used with stimuli. This hypothesis is examined experimentally in the following chapters.

There is at present no published evidence from other researchers about the fragmented memory effect. There is much evidence, however, on recognition performance when presentation time and interstimulus intervals are varied. These studies are reviewed below.

6.2 TEMPORAL PARAMETERS IN RECOGNITION MEMORY FOR PICTURES

Temporal parameters such as presentation time for each stimulus, the interstimulus interval (ISI) between stimuli, and delay between presentation of a list and the memory test have long been used as variables in the study of memory. These studies have important implications for an understanding of processes involved in remembering pictures. There is some evidence that memory processing for pictures increases when longer exposure times are used for each stimulus, but stops or diminishes during the ISI. This suggests that encoding is a process which requires a certain amount of time in order to extract information from the stimulus for storage in memory. This could be due to the time needed to operate a sequence of acts of attention, each of which has a maximum duration in time and a limit on the capacity of information which can be processed. Whatever the nature of the encoding process involved, it appears to operate while the stimulus is perceived directly and cannot continue when the stimulus is not physically present. However, there are some inconsistencies in findings of studies measuring recognition memory under variable conditions of stimulus exposure and delay. These must be resolved before the nature of the complexity of encoding processes can be understood.

There is a great deal of evidence that recognition of pictures increases when presentation time is lengthened from 60 milliseconds to five seconds. This suggests that encoding becomes more efficient when there is more time to process the stimulus physically present. Although an early study by Mooney [1958] failed to find any effect of presentation time on recognition of shapes, a later study with improvements in methodology [Mooney, 1960] did find that recognition of shapes presented for five seconds was superior to recognition of shapes presented for 70 milliseconds. Potter and Levy [1969] also found that recognition accuracy of coloured photographs of scenes increased significantly when presentation time was increased from 125 milliseconds to two seconds. Over the seven exposure times used, there was a positive relationship between the proportion of hits and exposure time. Similarly, Franken and Davis [1975] found that photographs presented for one second were recognized correctly more often than those presented for 100 milliseconds; and Hines [1975] found increasing recognition accuracy over exposure times varying from 150 milliseconds to one second. Using longer presentation times and a constant ISI of five seconds, Madigan and Rouse [1974] found that recognition of drawings in their correct orientation, as opposed to recognition of distractors which were reversals of the originals, was greater when items were exposed for four seconds rather than one second. All these studies provide firm support for the hypothesis that information which can be extracted from a picture for later retrieval (i.e., encoding) is a function of the amount of time available to process the stimulus. However, these studies varied only the exposure time for each stimulus. They did not vary the time after stimulus had been removed in which encoding could be continuing. Since most of the studies used an ISI of constant duration, there is the confounding factor that the subjects have longer "total time" to process pictures given longer stimulus exposures. Perhaps the total length of time available to process each stimulus (i.e., the presentation time and ISI) can be used together to improve stimulus processing.

Some studies have varied both presentation time and ISI in examining recognition of pictures, which allows a partialling out of effects of stimulus exposure and length of time available for encoding. A well-known study in the area by Shaffer and Shiffrin [1972] found significantly more positive confidence ratings in recognition of photographs as presentation time increased between 200 milliseconds and four seconds. Unfortunately, they did not report whether recognition accuracy increased significantly with presentation time, although they claimed that confidence reflected subjects' accuracy. Between ISI's of one and four seconds, however, there was no change in confidence ratings. This was thought to support the notion that pictures are not rehearsed or actively encoded after they have been exposed, but are processed completely in the presentation time. The Shaffer and Shiffrin results have been replicated with shorter exposure times and ISI's using recognition accuracy measures [Lutz and Scheirer, 1974]. The procedure was similar in this experiment, except that subjects viewed a random visual noise pattern during the ISI, whereas Shaffer and Shiffrin presented a darkened screen between items. The noise field should have prevented formation of icons or masking effects during the ISI, but it may also have disrupted normal encoding processes which are carried out after stimulus offset.

There is some evidence that a masking stimulus in the ISI disrupts normal processing. Loftus and Bell [1975] presented a group of subjects with a blank four second interval between each of 60 stimuli shown, while another group received one second of visual noise immediately after each stimulus, followed by three seconds of the blank field. Exposure times were varied from 60 to 500 milliseconds, and recognition accuracy increased with longer times. Performance under the masked presentation was significantly inferior to performance under the unmasked condition. Loftus and Bell did not examine ISI's of different durations, but their results suggest that a mask does truncate the visual processing which occurs in normal viewing.

In opposition to Lutz and Scheirer's [1974] findings, Tversky and Sherman [1975] found an improvement in recognition accuracy as ISI's increased from 1.5 seconds to three seconds as well as when presentation time was increased from 250 milliseconds to two seconds. This suggests that encoding can continue after stimulus offset. Tversky and Sherman noted that their data were consistent with the notion that total time is the critical variable in determining increasing accuracy over longer presentation times and ISI's. They did not offer evidence to suggest whether the processing which is carried out after stimulus offset is as beneficial for encoding as the processing carried out during picture presentation. In other words, they did not examine the question of whether increasing the length of presentation time has the same effect on memory as increasing the ISI. Obviously, the person cannot extract new information from a picture once it has been removed, unless he is acting on some sensory representation, such as an icon. The encoding processes carried out after stimulus offset must eventually diminish with increasing delay after offset, since the person is processing his own memory for the picture rather than a stimulus. Perhaps it is wise to distinguish between *encoding* processes, which occur while the subject is attending to a stimulus, and similar processes which occur after stimulus offset, sometimes called *rehearsal*. In both cases, the subject is attempting to organize and store information so that it can be retrieved later, but rehearsal may rely more on memory of the stimulus than encoding. A critical experiment in this area would examine effects of presentation time, ISI and total time on recognition of pictures. Such an experiment is carried out in Chapter Seven.

Tversky and Sherman [1975] criticized Shaffer and Shiffrin's [1972] results because of the ease of the recognition task. This is often a problem in study of recognition using heterogeneous distractors, as was discussed earlier. Tversky and Sherman used drawings as their stimuli, with homogeneous pictures as the distractors. Both studies used within-subject designs, but a methodological advantage of Tversky and Sherman's [1975] study was that temporal conditions were not presented in a random order. Instead, they were blocked, so that subjects could develop strategies for dealing with the amount of time given for each picture. This may have been one reason subjects in the Tversky and Sherman experiment could use the ISI to their advantage while subjects in the Shaffer and Shiffrin experiment could not.

The issue of effects of longer ISI's on memory is clouded by experiments in which the subject is presented with a single stimulus and is tested with a single test item after a period of time. These studies demonstrate that memory begins to decay as soon as the stimulus is removed. Hence, no rehearsal or continuation of encoding processes begun while the picture was present can be occurring. Posner and his colleagues [e.g., Posner, Boies, Eichelman and Taylor, 1969] have developed a recognition reaction time paradigm in which subjects are presented with a stimulus for a specified exposure. Then after a certain interval a recognition test item is presented and the subject makes a "yes" or "no" recognition response to this item. There is no ISI in this case, since the interval is really the delay between presentation of the item and recognition testing. Reaction time, the latency between onset of the test item and the subject's response, is measured. In this paradigm, recognition accuracy is very high, and often the experiment is controlled so that all subjects respond accurately. Phillips and Baddeley [1971] attempted to extend Posner's work with letters to a study employing five-by-five matrices with randomly filled squares, in order to study recognition reaction time in difficult-to-verbalize stimuli. All stimuli were exposed for 500 milliseconds with delay intervals of 300 milliseconds to nine seconds before presentation of the next matrix. The test matrix was either identical or differed from the original in having one square more or one square less filled in. Percentage correct recognition decreased significantly over the three to nine second interval. Mean reaction time also increased with delay for both correct and incorrect responses. In this case, during the delay the subject was looking at a masking field of randomly filled squares of a 10-by-10 matrix, a pattern which would be interfering for the icon of the stimulus, so that the interval was really filled with an interpolated task. This study suggests that forgetting begins almost as soon as the pictorial stimulus is removed and continues for at least nine seconds.

This finding has also been replicated by Nickerson [1972], who demonstrated that recognition accuracy decreased when the delay between stimulus and test items was increased over intervals from zero to eight seconds. Nickerson also used random matrices, although they were seven-by-five dot patterns, and presented a seven-by-five masking matrix of dots in the delay intervals. In both these recognition reaction time studies [Nickerson, 1972; Phillips and Baddeley, 1971], it may be suggested that the masking stimulus interferes with processing in the delay interval.

Another recognition reaction time study did not employ a mask in the delay interval and found that memory need not decay rapidly with the length of time after the stimulus has been removed. Using "free form" closed-line drawings, Cermak [1971] employed the Posner paradigm to examine recognition accuracy for shapes presented for five seconds and tested with another item after delay intervals from 1.5 to 20 seconds. No masking stimulus was used, and subjects were made aware of the delay interval to be used in each block of stimuli presented in each delay condition in the within-subject design. Accuracy decreased significantly over delay, with a significant drop from 1.5 seconds' to four seconds' delay, but not from four seconds' to 20 seconds' delay, although performance was still significantly better than chance at 20 seconds' delay. Thus, information available after four seconds was still available 20 seconds later. These results seem to conflict with those of Phillips and Baddeley [1971] and Nickerson [1972]. Since Cermak used an unfilled delay interval instead of a masking pattern, subjects may indeed have used the delay to their advantage with some sort of rehearsal mechanism or remembering strategy. Alternatively, an interfering activity (e.g., examining a mask) may interact with normal forgetting to increase the decay function.

The introduction of a mask in the interval after a picture has been shown is an artificial procedure which appears to decrease memory performance, and perhaps to disrupt pictorial processing entirely. Cermak's study is the only one using the Posner paradigm which avoids the use of a mask. On the other hand, the studies by Phillips and Baddeley [1971] and Nickerson [1972] examine the amount of information which can be extracted from a pattern during the time it is physically present. Though this may be of interest in itself, it offers no evidence about the amount of information which can be remembered about a picture under normal viewing circumstances. None of the studies employing masks in the ISI's between items in a list of pictures [Lutz and Scheirer, 1974] or in the interval between stimulus and test in a recognition reaction time task [Nickerson, 1972; Phillips and Baddeley, 1971] have found beneficial effects for memory of time after stimulus offset. However, studies employing no masks have found that lengthening the time after stimulus offset before the presentation of any other picture either does not hinder memory [Cermak, 1971; Shaffer and Shiffrin, 1972] or actually aids recognition [Tversky and Sherman, 1975]. This suggests that under normal viewing conditions, when stimuli are not followed by masks meant to interfere with their icons, subjects can continue to extract information from the icon, better organize what is remembered about the stimulus as it is being registered in memory or ensure that the information is in a retrievable form.

Further support for the view that subjects may be able to use the ISI for extra processing of the stimulus comes from a study of the serial position curve in recognition of pictures. Hines [1975] found a noticeable recency effect in recognition of four sequentially presented random shapes when testing was carried out immediately, after 30 seconds of delay, or even after 30 seconds of an interfering task. Hines [1975, p.638] suggested that, for his lists of shapes in which no ISI was used (and hence each successive shape masked the previous one):

> "It appears probable that the last stimulus continues to be processed into (long term memory) for a brief period following actual offset."

The effect of masking due to no ISI in this experiment appears to have been severe in reducing recognition of the shapes, truncating valuable processing time, so that only the last picture in the list in Hines' experiment, which was not masked by its successor, showed less decline in memory over time. Of course, the last item in the list has the shortest delay between presentation of the list and the test. However, if this were the reason for its better memorability and not the fact that it was the only unmasked stimulus, there should have been a positive linear function relating memory accuracy to the position of each item in the list, rather than the marked recency effect found.

In conclusion, it is not exactly clear how temporal parameters affect memory for pictures. The studies reviewed have consistently shown that recognition increases as presentation time increases from 60 milliseconds to five seconds. There is probably an optimum presentation time for each type of stimulus. This finding suggests that with more time for viewing the stimulus, subjects can use pictorial information to greater advantage. They can encode more of the picture, since they have time to make more detailed attentional scans, and subjects can employ their memory strategies to aid later retrieval of the information encoded. Shorter presentation times may curtail any of these processes.

It is not clear what happens to pictorial processing once the picture has been removed. Traditional recognition studies present conflicting results. Two studies have indicated that varying the interval between pictures in a list from 200 milliseconds to four seconds has no effect, but one study did not report accuracy data [Shaffer and Shiffrin, 1972] and the other filled the delay interval with a mask [Lutz and Scheirer, 1974]. A later study [Tversky and Sherman, 1975] found that recognition increased as ISI was lengthened from 1.5 to three seconds. This study did not employ a mask in the ISI. The findings are not clarified by examining evidence based on another paradigm, the recognition reaction time method, since there is no comparable ISI. Rather, each to-be-remembered picture is followed by a single test item, so that the interval is more appropriately labelled a delay interval than an ISI. Delay has consistently been found to decrease recognition of pictures (see Table 1 and Experiment 1 in Chapter Three). Since these recognition reaction time studies do not define delay differently to ISI, they are not comparable to studies employing lists and perhaps their findings should not be taken as evidence that memory declines over the ISI in a list. In a list recognition paradigm, ISI is carefully defined, while delay is variable, since the time between presentation of any stimulus and its test item is usually randomized. To separate the effects of these two temporal variables, it may be wise to study their differential effects in the same experiment.

On the basis of Tversky and Sherman's [1975] findings, it appears that under normal viewing conditions, the subject can use the ISI to continue active stimulus processing. However, a beneficial effect of longer ISI's could also be due to passive processes involved in consolidating a memory trace. The memory system may operate most efficiently when information is not absorbed too rapidly. A longer ISI could provide a pause in the sequential attention scanning process essential to the encoding of an event. Thus the system may need to "assimilate" information from one stimulus before proceeding to the next. The longer ISI also allows each event to be registered in a more unique temporal context, which may aid the establishment and retrieval of memory traces.

However, if the advantage of a longer ISI is in providing a "rest" in the system for passive consolidation of a memory trace, it seems inconsistent that a masking stimulus could disturb this process, as it clearly does, unless consolidation requires the unhindered formation of an icon. Since icons are not a normal part of remembering stimuli in everyday activities (since each perception is "masked" to some degree by the person's continuous visual scanning), the formation of an icon would not seem to be the only necessary process truncated by the highly interfering masks of some pictorial experiments. The most likely explanation seems to be that if the time in the ISI can be used to benefit memory, it is mainly due to some active process which begins while the stimulus is present and continues for a short time after stimulus offset. If this were true, the next problem would be to determine whether the process changes once the stimulus has been removed.

6.3 HYPOTHESES

The research reviewed in the last section suggests that subjects encode more information from pictures the longer they are available during presentation. This implies that with decreasing presentation time, there will be a greater likelihood that only part of a stimulus will be remembered. The presence of the fragmented memory effect suggests that this part can be remembered on its own, without other information from the picture, and that the demands of a recognition memory task force the subject to try to place this part into the context of a whole picture. It is uncertain, though, whether subjects use the fragmented parts in memory to confabulate a whole picture during storage or retrieval, or whether the subject actually knows that he only remembers fragments and normally would not attempt to construct the parts into a plausible whole. One assumption of the exclusion set recognition method is that the subject's most realistic reponse in recognition is to reject a picture with any detail which is contradictory to what is remembered. Following from this, responses which appear to reflect the fragmented memory effect should be uncertain acceptances of any test stimuli which do not contain items inconsistent with the incomplete memories. One purpose of the recognition experiments to follow was to examine whether memory for both parts of a shape as well as their combinations increases with longer time to process the stimulus, i.e., with longer presentation times. This has not been examined by previous researchers.

The experiments to follow also examine the effect of longer ISI's on recognition of parts of pictures and their combinations. Since the evidence reviewed concerning this issue is unclear, one purpose of the experiments was to determine whether there are any beneficial effects on memory of increasing or decreasing the time elapsed between items in a list of shapes when no masking stimulus is employed. It was hypothesized that an increase in ISI would aid recognition of whole shapes.

A final purpose of these experiments was to demonstrate the fragmented memory effect. Since the effect may be due to the subject's lack of time to encode each part of the picture, it was hypothesized that the fragmented memory effect would increase with shorter presentation times. If consolidation or some active rehearsal process is necessary to the process of encoding events, so that remembered parts of a picture can be stored in a way that enables them to be retrieved as a whole, then it seems likely that the separation of each event in time would provide a better temporal context in which an event could be encoded for later retrieval. Thus it was also hypothesized that the fragmented memory effect will decrease with longer ISI's, but only up to a certain point when decay starts to affect memory.

CHAPTER SEVEN

INVESTIGATION OF TEMPORAL PARAMETERS IN RECOGNITION OF PICTURES

7.1 THE METHOD FOR TESTING RECOGNITION OF PARTS AND WHOLES

In order to test the effects of variable presentation time and ISI on memory, a new method of recognition was needed to measure memory for the parts of a stimulus as well as the whole configuration. Two-part abstract shapes were chosen as stimuli, to allow greater control over their physical properties and to minimize labelling, which seems to be evoked as a mnemonic strategy in the experimental situation. The shapes, described in the last chapter, were distinctive configurations which had more visual structure (due to their symmetry) than random shapes. This seemed to make them easier for subjects to remember.

Usually, stimuli were presented in lists of three, and then two of these were tested in the recognition task. The parts (inside figures and outside polygons) of the remaining stimulus were used in creating recognition distractors. Memory for each stimulus was tested by requiring subjects to select one of a set of six, simultaneously presented items. Simultaneous presentation of test items was necessary because of the difficulty of the task, since subjects forgot the stimuli very quickly. The six items in each test set comprised:

(1) an original shape,

(2) a shape made of one part of the original in combination with a part from another shape in the stimulus set (which was not tested),

(3) the inside part of the original with a completely new outside,

(4) the outside part of the original with a new inside,

(5) a combination of the two new parts to make a whole, and

(6) a combination of the part taken from the untested stimulus with one of the new parts.

It will be convenient to refer to these items in the test sets by means of the following notation. Each picture can be designated as an "*xx*" item, in which the first letter denotes whether the inside figure was old (*o*) or new (*n*) and the second letter refers in similar manner to the outside figure. Thus, correct stimuli will be described as *oo* items and distractors as *nn* items. Incorrect combinations of old parts will be designated as *oo'* items. To test for the subject's memory of the outside parts only, an old inside was paired with a new outside (*on* items), while memory for inside figures only was tested with items pairing a new inside with an old outside (*no* items). Each test set contained these five types of items, plus an additional *no'* item to balance the number of occurrences of any one part in the test set so that no parts were too distinctive in the set. For the half test sets, *no'* and *oo'* items were used, and for the other half *o'n* and *o'o* items were used. An example of a test set is shown in Figure 11.

There is a problem with subjects' strategies in this testing method. The subject is required to guess between those items which he cannot reject, and there is no absolute certainty about the kind of memory that a choice of one of the six types of item would imply. For example, Table 17 shows the choices consistent with each of nine possible memories of the stimulus.[1] Ideally, it would be best to estimate the probabilities of each of the nine possible contents of subjects' memories (Table 17) using the observed frequencies of choosing each of the six alternatives in the test sets. Unfortunately, the probabilities associated with a true content of memory cannot be derived from the frequencies of response choices. The exception to this is the case of no memory. It can be supposed that where more than one response type is consistent with the subject's memory state, the subject chooses randomly between these alternatives. Then, the probability that the subject has remembered nothing (i.e., that the subject is guessing) can be esimated as six times the frequency of selection of the *nn* item. However, a

---

[1] This is shown only for sets in which *oo'* and *no'* items appear. An equivalent table can, of course, be drawn for sets with *o'o* and *o'n* distractors.

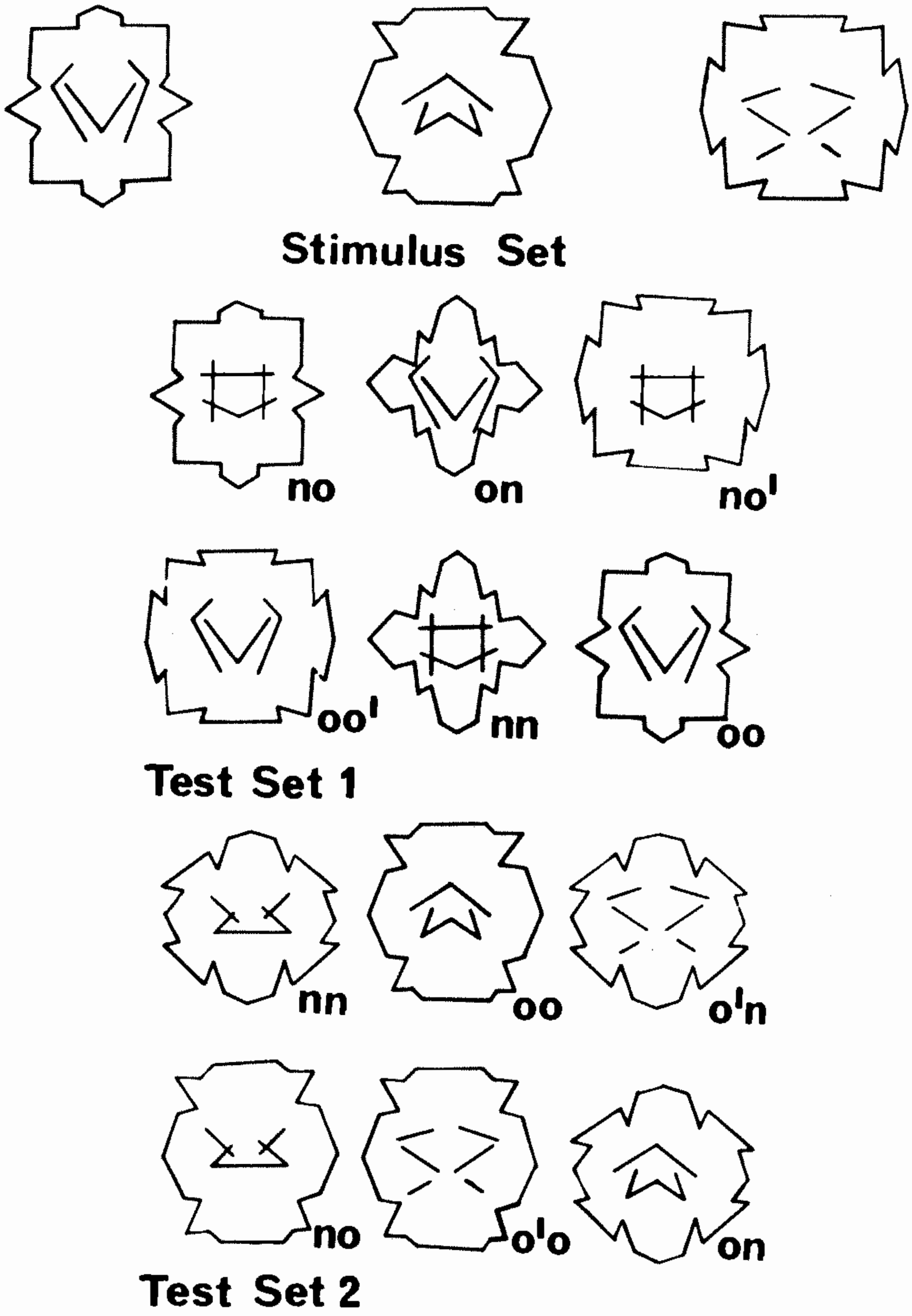


Figure 11. A stimulus set of three shapes and the two recognition sets which were used to test memory for the first two stimuli (Experiment 5). The third stimulus (top, right) was not tested, and its parts were used to make the *oo'* and *o'o* distractors. Each item in the test sets is labelled. These labels were not presented during testing.

Table 17

Possible contents of memory implied by a choice of each recognition item.

Memory Content	Possible Choices of Items in the Test Set					
	oo	*oo'*	*on*	*no*	*no'*	*nn*
1. the whole configuration	x					
2. the two correct parts without the whole	x					
3. the outside only	x			x		
4. the inside only	x	x	x			
5. the outside of the non-tested item only		x			x	
6. three old parts	x	x				
7. two outsides only	x	x		x	x	
8. the correct inside and the outside of the non-tested stimulus only		x				
9. nothing	x	x	x	x	x	x

variety of partial memory states can contribute to all other response selections, and these cannot be isolated from each other.

It is possible to obtain a measure of the tendency towards accurate memory responding by examining the *oo* rate, treating the other items in the recognition test as ordinary distractors. If the *nn* rates for various experimental groups differ, then it may be possible to base a guessing correction for the *oo* scores on the corresponding *nn* rates. However, like all traditional recognition measures, this would not differentiate between *oo* responses based on perfect or partial memory accuracy.

The fragmented memory effect can be measured in this paradigm by comparing proportions of *oo* and *oo'* responses. In both cases, the subjects have remembered parts of the stimuli seen. The subject who has remembered only the parts of the stimuli and who has failed to encode the shapes as wholes should not be able to differentiate between *oo* and *oo'* items and should distribute his responses randomly between these items in guessing. On the other hand, the subject who has an accurate memory for the whole configuration should choose the *oo* item more often than the *oo'* item. As a measure of fragmented memory, a "correct combination of parts" score (CCP) can be calculated for each subject using the formula

$$\text{CCP} = \frac{\text{Frequency } (oo) - \text{Frequency } (oo')}{\text{Frequency } (oo + oo')} .$$

In other words, given the occasions on which subjects have chosen at least two correct parts (*oo* and *oo'* items), the proportion of these which indicate that memory of the whole is lost (*oo'* responses) is subtracted from the proportion of correct (*oo*) responses. A positive CCP score indicates that the subjects are accurate in choosing combinations and can discriminate between *oo* and *oo'* items. A null score indicates that subjects have memory for parts but no information about their relationships. The denominator of the CCP score is itself a meaningful score, since the total frequency of *oo* and *oo'* choices can be seen as a measure of the subject's memory for correct parts of pictures seen. This score will be referred to as the "total parts" (TP) score.

To see whether subjects are biased towards remembering only an inside part of an outside part of the shapes, total *on* and *o'n* responses are compared with *no* and *no'* responses. The guessing rate for each of these responses is higher than that for other distractor items, since there are either two correct insides or two correct outsides paired with a new part in each distractor set. Subjects have a 25% chance of choosing an *on* or *no* item at random, but only a 17% chance of choosing an *oo* or *oo'* (or *nn*) item at random. For this reason, *oo* and *oo'* scores are not compared directly with *on* or *no* scores in statistical analysis.

7.2 OVERVIEW OF THE EXPERIMENTS

In the experiments described in this chapter, presentation time and ISI were varied to determine whether they affect recognition memory for shapes. It was hypothesized that longer presentation times and ISI's would lead to a higher hit rate (more *OO* responses) and less effect of fragmented memory (higher CCP scores). In Experiment 4, fast presentation times and ISI's of 0.5 seconds and two seconds were varied in all four possible combinations. Because of methodological problems in this experiment, the testing procedure was changed, and Experiment 7 was carried out as a partial replication. Experiment 7 is reported last in order to illustrate the chronological order in which the experiments were carried out and the evolution of the methodology. Only presentation time was varied in Experiment 5. This was the only experiment employing a within-subject design. Subjects' memories were compared under presentation times of two and five seconds when ISI was held constant at 1.5 seconds. Finally, in Experiment 6, the only temporal variable manipulated was ISI. Subjects given 1.5 seconds ISI were compared in performance to subjects given seven seconds ISI, when presentation time was held constant at two seconds. Taken together, the experiments investigate recognition memory for shapes over presentation times from 0.5 seconds to five seconds and ISI's from 0.5 to seven seconds.

7.3 EXPERIMENT 4: A STUDY OF EFFECTS OF RAPID PRESENTATION TIMES AND ISI's ON RECOGNITION MEMORY

7.3.1 Method

Design

Four different experimental groups were tested, each being given stimulus lists with a different combination of presentation time and ISI. Both presentation time and ISI were varied between two durations, 0.5 and two seconds. The table below represents the design graphically.

		Presentation Time (seconds)	
		0.5	2
ISI (seconds)	0.5	Group 0.5/0.5	Group 2/0.5
	2	Group 0.5/2	Group 2/2

Subjects

Thirty-six first-year Psychology students at the A.N.U. participated in the experiment. Each of the four experimental groups had nine subjects. Volunteers were paid $2.00 for their participation. There were five males and four females in all groups except the 2/2 group, which had four males and five females due to an error.

Materials

There was an attempt to make the pool from which stimuli were chosen as large as possible. For the outside parts, squares of paper were folded over twice and cut from corner to corner with six continuous straight lines to make 24-sided polygons. The shapes thus produced were opened out and traced onto paper. Interior shapes were made by drawing three line segments onto a folded piece of paper and tracing through to the other side to make symmetrical six-line-segment figures. Sixty different pictures were formed by choosing combinations of parts which were as distinctive as possible from each other.

Selection of Stimuli

Several variables must be controlled in any study using pictures as stimuli. The verbalizability of the pictures must be controlled, since amount of verbal encoding may affect the memory process. Ease of naming (codability) and richness of verbalization are both important. For shapes such as those used in the present experiment, similarity of the pictures, their complexity as line drawings and their visual structural coherence or "articulation" are all important. Familiarity of the shapes is another relevant variable, but this is a difficult concept to define operationally.

For stimuli in the present experiment, it was decided that the most important variables to be controlled were verbalizability and the overall structural coherence of the total picture. The latter variable was chosen because some of the pictures seemed to have redundancies between their interior design and the background outline which would allow them to be more easily perceived as a coordinated whole and not as two unrelated stimuli. This might influence the fragmented memory effect.

A ratings study was carried out to choose the pictures lowest in verbalizability and ease with which the parts of the pictures could be linked together perceptually. Twenty judges were given the stack of 60 pictures, which were shuffled before each presentation. Each picture had a number in its bottom left corner which judges used to identify the pictures. Judges were asked to make two kinds of judgments of each picture. For the ratings of unity of the parts of the picture, judges were told,

> "For each picture, determine how easy it is to link each part (the inside and the outside of the drawing) together. Some pictures may have parts that 'go together' better than others. This may be a useful strategy in memory for linking the parts together so that they can be remembered as a whole."

The judges were asked to rate each shape on a seven-point scale from "no relationship between the parts" (1) to "an obvious relationship between the parts" (7).

For measuring verbal meaningfulness of the pictures, judges were told,

> "For each picture, in the large box provided, write down a list of words which come into your mind to describe it or any parts of it. Leave the space blank if you can't think of anything."

Judges were given sheets of paper marked off into squares, with a small box in each square for the number of each picture rated. Before starting, they were asked to look through all the pictures to get an idea of the range of materials to be used.

Verbal responding to each shape was scored by the frequency of words used, omitting superfluous conjunctions and articles. The number of these associations ranged from 1.65 for one shape to 4.60 for another. The 25 items having the lowest scores on both ratings were chosen as stimuli, with 15 randomly chosen to be the stimuli and 10 randomly chosen to be the *nn* distractor items.

Stimuli

The 15 stimuli were photographed on 16 mm moving film in a random order. Shapes were presented as negatives to minimize brightness. One moving film was made for each experimental condition, so that presentation times and ISI's could be set at either 0.5 seconds (12 frames of film) or two seconds (48 frames of film).

Test sets were made for each list of stimuli according to the procedure outlined in section 7.1. Of the 15 stimuli presented, only 10, chosen at random, were to be tested in recognition. The remaining five shapes were a source of parts for the distractor items. Each of the 10 *nn* items was used in a different recognition test set, and was paired randomly with an *oo* item. Parts of *oo* and *nn* shapes were put together to make the *on* and *no* distractors. In each list of three shapes, the stimulus chosen at random to be the untested item was separated into its two constituent parts. The inside part was chosen to make the *o'o* and *o'n* distractors, in one test set, while the outside part was chosen for the *oo'* and *no'* distractors in the second test set. Positioning of the six items in each test set was randomly determined for each two-by-three array. The test sets were photographed and reduced to 35 mm negative slides. The stimuli are shown in Appendix 4.

A practice set was constructed using shapes similar to the test pictures, but different in line content (see Appendix 4). A test set was made for these using the method described above. The practice shapes and their test set were presented as 35 mm negative slides.

Apparatus

Stimuli were shown with a 16 mm Kodak moving film projector. The images projected onto the centre of a rear projection screen. During recognition testing, slides were presented with a Kodak Carousel slide projector, Model S-RA, onto the same area of the screen. The screen was placed approximately one metre from the subjects so that projected images subtended a visual angle of 14°.

Procedure

Pilot testing determined that it was impossible to show subjects all 15 stimuli in a row, since recognition performance was then very low. Subjects were shown three lists of three pictures followed by one list of six pictures. Subjects were told,

> "I am going to show you three pictures one after the other. Look at each one carefully. They will be presented very quickly. Then afterwards, I want to see if you can recognize the original pictures by picking them out from some similar pictures."

Each subject was shown the four practice pictures one after the other, with approximately two seconds presentation time and 1.5 seconds ISI. Then the test set was shown. The experimenter asked the subject if he could choose one of the items in the "memory set" as "being most like one of the original" shapes. Subjects were asked to make their answers on a response sheet resembling the form below by:

		Rating		
Number	Set	Inside	Outside	Whole
	● ● ● ● ● ●			

(1) inserting the number of test set shown at the bottom left corner of the slide under "Number",

(2) circling a dot in the two-by-three dot matrix under "Set" to denote the position of the item on the test slide which subjects chose as the remembered item, and

(3) giving a confidence rating about the "inside" or "outside" part alone of the item in the test set chosen as well as about the combination of these parts (the "whole").

Subjects were given a card with the following three confidence ratings printed on it:

(1) I am positive it is the original.

(2) I am fairly sure it is the original.

(3) I think it could be the original.

Once it was determined that subjects had understood the rating scale and had rated the insides, outsides and wholes of the items chosen, they were told that the task would begin and that three pictures would be presented in rapid succession, followed by a memory slide. The subjects were shown the film clip appropriate for their experimental group containing the first three stimuli. Then the first test set was shown. Timing of these slides was self-paced, and subjects were asked to inform the experimenter as soon as they finished the first set so that the next set could be presented. Recognition testing proceeded in this way for the second and third lists of stimuli. Before the last list, subjects were told that it would contain six pictures and would be tested by four successive recognition sets. It was originally intended to have two random orders of each stimulus list by running the film backward as well as forward, but the use of a longer list at the end made this impossible, since it may have been more difficult. The order of the stimuli in each list, the order of the test slides following each list and the order of lists were randomized for each subject.

At the end of the recognition procedure, a questionnaire was administered orally to subjects. They were asked the following questions:

(1) Did you find anything difficult?

(2) Did the practice set prepare you for the test picture?

(3) Did you use any particular strategy in trying to remember the pictures?

(4) Did you find yourself concentrating more on the inside, the outside or did you see the whole thing globally?

(5) Did you try to name anything as it came up?

7.3.2 Results

Means and standard deviations of proportions of responses to *oo*, *oo'* (including *o'o*), *on* (including *o'n*), *no* (including *no'*), and *nn* items are shown in Table 18 for each presentation group. Means and standard deviations of CCP and TP scores are also shown for each group. Proportions of both the hit rates and the TP scores are higher than would have been expected by chance (which would be 0.17 and 0.33, respectively), demonstrating that subjects were able to remember some of the stimuli. The actual guessing rate, as determined by multiplying the *nn* rate by six, ranged from 12% to 24%. The CCP scores are all fairly close to zero, suggesting that the fragmented memory effect was quite large in the experiment.

Table 18

Means and standard deviations of proportions
of responses to all items in the test sets, TP and CCP scores
for each experimental group in Experiment 4.

Score		Group			
		0.5/0.5	0.5/2	2/0.5	2/2
oo	$\bar{X}$	0.2556	0.2333	0.3000	0.2667
	S.D.	0.0153	0.0175	0.0350	0.0100
oo'	$\bar{X}$	0.3111	0.2444	0.3111	0.3444
	S.D.	0.1166	0.0882	0.1833	0.1333
on	$\bar{X}$	0.2222	0.3111	0.2333	0.2889
	S.D.	0.1204	0.1054	0.0866	0.1054
no	$\bar{X}$	0.1667	0.1556	0.1222	0.0667
	S.D.	0.1658	0.1237	0.0975	0.0707
nn	$\bar{X}$	0.0444	0.0444	0.0333	0.0222
	S.D.	0.0529	0.0529	0.0500	0.0447
TP (*oo* + *oo'*)	$\bar{X}$	0.5667	0.4556	0.6111	0.6222
	S.D.	0.1871	0.1740	0.1167	0.1481
CCP $\left(\frac{oo - oo'}{oo + oo'}\right)$	$\bar{X}$	-0.1084	-0.0782	-0.0181	-0.1119
	S.D.	0.0651	0.1525	0.3533	0.0829

Hit Rate Analysis

To determine whether subjects' memory accuracy was affected by the presentation conditions, a two-way analysis of variance was carried out on *oo* scores with two levels (0.5 and two seconds) for each of the random factors of Presentation Time and ISI. The summary table is

Table 20

Summary of the analysis of variance for CCP scores: Experiment 4.

Source	SS	df	MS	F	p
ISI	0.0073	1	0.0073	0.0447	NS
Presentation Time	0.0091	1	0.0091	0.0557	NS
ISI × Presentation Time	0.0346	1	0.0346	0.2118	NS
Within Cells	5.2297	32	0.1634		
Total	5.2807	35			

respectively, were found. None of these is significantly different from zero using the 5% significance level of a two-tailed test. Subjects could not respond differentially to *oo* and *oo'* items. Thus, the fragmented memory effect was an important error factor in recognition responses of subjects.

Tendency to Remember only One Part of a Shape

Inside parts of the shapes appeared to be easier to remember than outside parts in this experiment. Subjects more often chose incorrect (new) outsides than insides, with mean frequencies of 2.8 *on* responses and 1.6 *no* responses over all groups. There were significantly more *on* than *no* responses when scores for all groups were combined ($t = 3.813$, $df = 35$, $p < .002$ for a two-tailed test). This finding was supported by the questionnaire data. A majority of subjects (86%) reported that they concentrated mainly on the inside parts or found them easier to remember, while only 6% of subjects concentrated on the outsides and 8% concentrated on one part or the other in alternating fashion.

Sex Differences

To test for sex differences, mean hit rates of male and female subjects were compared. Male subjects averaged 27% *oo* scores (S.D. = 13%) and female subjects averaged 26% *oo* scores (S.D. = 15%). No significant sex differences were found ($t = 0.3208$, $df = 15$, $p > .05$ for a two-tailed test). Because of uneven numbers of males and females in some groups, three males and one female were rejected randomly from the analysis so that all groups included four males and four females for deriving the pooled mean scores.

Confidence Ratings

Mean confidence ratings for insides, outsides and wholes of the shapes for each item type chosen in recognition, and their standard deviations, are shown in Table 21, for each experimental group. Subjects were most confident of their memory for insides and least confident about *nn* choices, as expected. There was little difference in confidence ratings for *oo* and *oo'* items, but insides of *oo'* items were rated with less confidence than insides of *oo* items. The new outsides or insides of *on* and *no* items were given lower confidence ratings than their old counterparts in the shapes, suggesting that subjects were aware of the occasions on which they were guessing.

Questionnaires

Subjects' answers to the post-recognition questionnaire revealed that 44% of subjects in the groups given 0.5 seconds presentation time and 67% of subjects in the groups given two seconds presentation time verbalized about the stimuli in some way when they appeared. Only subjects in groups given the longer presentation time mentioned that they used a memory strategy attempting to link the inside and outside parts together: one-third of subjects in the 2/0.5 group reported this, while only one subject in the 2/2 group reported it.

7.3.3 Discussion

The fragmented memory effect was clearly demonstrated in this study. Subjects in all groups were unable to distinguish between original pictures (*oo* items) and incorrect combinations of old parts from different stimuli (*oo'* items). The low rate for choosing

Table 21

Means and standard deviations of confidence ratings of insides (I), outsides (O) and wholes (W) for each item type in each experimental group in Experiment 4.

Group		Response Type														
		oo			*oo'/o'o*			*on/o'n*			*no/no'*			*nn*		
		I	O	W	I	O	W	I	O	W	I	O	W	I	O	W
0.5/0.5	$\bar{X}$	1.91	2.44	2.52	1.46	2.11	2.29	1.75	2.50	2.65	1.93	2.20	2.60	2.75	2.00	2.50
	S.D.	0.67	0.73	0.59	0.58	0.74	0.76	0.72	0.61	0.49	0.80	0.78	0.63	0.50	1.16	0.58
	frequency		23			28			20			15			4	
0.5/2	$\bar{X}$	1.82	2.00	2.27	1.77	2.14	2.09	1.50	2.21	2.57	2.36	2.14	2.00	2.75	2.50	3.00
	S.D.	0.85	0.76	0.77	0.81	0.71	0.61	0.69	0.79	0.57	0.63	0.87	0.88	0.50	1.00	0.00
	frequency		22			22			28			14			4	
2/0.5	$\bar{X}$	1.67	2.04	2.15	1.29	2.11	2.25	1.86	2.10	2.33	1.91	2.00	2.64	1.67	2.67	2.67
	S.D.	0.79	0.76	0.72	0.46	0.57	0.65	0.73	0.63	0.58	0.70	0.63	0.67	0.58	0.58	0.58
	frequency		27			28			21			11			3	
2/2	$\bar{X}$	1.44	1.56	1.72	1.58	1.87	2.19	1.58	2.08	2.35	2.33	2.00	2.17	2.00	2.00	2.50
	S.D.	0.71	0.65	0.79	0.72	0.76	0.75	0.70	0.69	0.63	0.82	1.10	0.75	0.00	1.41	0.71
	frequency		25			31			26			6			2	
Total	$\bar{X}$	1.71	2.00	2.17	1.53	2.06	2.21	1.67	2.22	2.48	2.13	2.09	2.35	2.29	2.29	2.67

completely new items (2% to 4%) shows that the results were not due to the overall difficulty of the task, since subjects responding randomly would have chosen the *nn* items on 17% of occasions. The frequency of choosing either *oo* or *oo'* items was above chance, which also demonstrates that subjects had some correct information about parts of the pictures they had seen, even if they were unable to remember the whole configurations.

It is interesting to note that accuracy did not change with increasing presentation time or ISI. Adequate encoding may require more time than the maximum of two seconds presentation time given for each shape in the present experiment. However, the studies reviewed in section 6.2 showed that presentation time has consistently been found to affect recognition of a variety of pictures. The results of the present experiment fail to replicate this well-known trend. This may have been due to methodological difficulties in the present experiment, such as high confusability between items in the stimulus and test lists. Thus, the results of the present experiment should be handled with caution. The three experiments which follow were carried out to test the generality of the present results.

There was a decreasing frequency of *no* responding with longer times for each stimulus, and a slight increase in *on* responding. The *on* responses were greater than the *no* responses in every group. Since accuracy did not differ in the four groups, as demonstrated by similar *oo* and *nn* scores as well as by similar combined *on* and *no* scores, it appears as though subjects were prejudiced towards attending to the insides with longer total time per stimulus. Perhaps this merely demonstrates a greater consistency among subjects in choosing an efficient attentional strategy when they have time to implement one. The results suggest that often subjects had time to attend to only one part of each picture.

The attempt to control stimuli in the present experiment by minimizing the verbal codability and associations between parts of the shapes using ratings did not seem to control the stimuli adequately. With only 0.5 seconds per picture, nearly half the subjects managed to verbalize occasionally about the stimuli during presentation. It may be impossible to find stimuli which subjects are not able to encode verbally in some way. The tendency to be able to see the pictures as unified, coherent wholes was suggested spontaneously only by the subjects who had been given longer exposures for each stimulus. It appears that stimuli do not have an intrinsic "unity of parts" in the way measured. Perhaps subjects can perceive pictures as wholes more easily with more time. The same may be true of verbalizability. Since temporal factors were manipulated in the experiment, verbalizability probably varied more with time allowed for each picture rather than with any intrinsic codability of the stimuli. It could also be that the entire stimulus pool was so homogeneous on these ratings that the attempt to choose those most similar as stimuli was unnecessary.

Another problem with the present experiment was that the moving film appeared to make the task more difficult than anticipated. There was noticeable flickering of the projected images, which may have been emphasized by their high contrast. Since the use of moving film makes it impossible to randomize the order of presentation of stimuli, it is much less satisfactory than presentation of slides.

The failure of the present experiment to find an increase in the *oo* rate with longer presentation times suggests that the experiment was inadequate or that the particular stimuli used were unusual in their patterns of memorability. The finding of a large fragmented memory effect may have been produced artificially by the confusability of the stimuli. In the experiment to follow, the confusability of shapes seen together was minimized to try to control this unsatisfactory aspect of the testing method.

7.4 EXPERIMENT 5: THE EFFECT OF LENGTHENING PRESENTATION TIME ON RECOGNITION OF SHAPES

Experiment 4 did not find any increases in recognition accuracy or decreases in the fragmented memory effect with longer presentation times or ISI's. However, the study failed to replicate a finding from previous studies that recognition accuracy increases when presentation time is increased from 0.5 to two seconds [e.g., Lutz and

Scheirer, 1974]. This suggested that the recognition task may have been too difficult and that subjects found it impossible to remember the stimuli and to make a recognition choice between such confusable items.

In the present experiment there was no attempt to control the verbal codability, structural unity or any other physical characteristic of the stimuli. It was decided that the stimuli were controlled for their homogeneity adequately by the strict rules used to generate the shapes. The most important aspect of the recognition task, the similarity between distractors and originals in the test sets (see Chapter Two), had not, however, been controlled adequately in Experiment 4. In the present experiment, judges were asked to choose sets of items which could be confused with each other. On the basis of this, items in sets were chosen to be as dissimilar to each other as possible.

To reduce the difficulty of the task, stimulus lists were shortened to three items for all sets in this experiment. Also, presentation time was re-examined using longer intervals, two seconds and five seconds, to see whether Experiment 4 failed to replicate the findings of previous studies because the shapes could not be perceived with brief exposures. Slides were used rather than moving film to overcome the disadvantages found with the latter. To minimize the high variance of scores in Experiment 4, a repeated-measures design was employed and more stimuli were tested in recognition.

7.4.1 Method

Subjects

Ten females and 10 males, chosen randomly from the pool of first-year Psychology students available, acted as subjects in the experiment. They were paid $1.00 for the 25-minute experiment.

Design

Two presentation times were used, two seconds and five seconds. All subjects received four trials of one condition followed by four trials of the other condition, in a counterbalanced order across subjects. The ISI was kept constant at 1.5 seconds. This was the shortest time possible using a single slide projector for the presentations.

Stimuli

The pool of 60 shapes created for Experiment 4 were used as the source of stimuli in this experiment. However, 20 pictures were omitted by several judges because they were very similar to other pictures in the pool. The 40 remaining shapes were presented on separate cards to 10 judges. Cards were displayed on a table in a five-by-eight matrix. The judges were told,

> "Put the pictures into piles. I am interested in finding out which pictures can be confused with each other in a memory task, so put pictures that are similar together. Make several piles and put at least four pictures in a pile."

It was decided not to use a totally free classification task, as pilot judgments had shown that some subjects will look only for pairs of similar pictures, or some other grouping which would not aid in selecting several nonconfusable pictures for each group of stimuli. The constraints of the sorting task were selected to facilitate the choice of least-confusable test sets.

Judges made two sorts of the pictures: once on the basis of the inside figure and once on the basis of the outside background. They were asked to make the first sort on the basis of the inside or outside part in a counterbalanced order. Pictures were reshuffled before each sorting. (Subjects were not given the inside and outside parts in separate presentations, since judgments about an isolated part may be different to judgments made about that part in its total visual context. Hence the judges were allowed to perceive the stimuli in the same way as subjects in the memory experiment.)

Two matrices were produced on the basis of the ratings, one for insides and one for outsides of shapes, showing the number of judges who had grouped each possible pair of items together as confusable. From this, it was possible to choose six sets of five pictures which no judges found confusable with each other. With the remaining 10 pictures, two sets of five were derived which were least confusable with each other. Only two judges found confusions in one set and for the last set, four judges found confusions between the insides of two *oo* items.

Sets of five shapes were chosen so that they were low in confusability. These sets provided three stimuli (randomly chosen) and two *nn* items to be used in the test sets for the stimuli. Of the three stimuli, two were randomly chosen to be tested. The remaining shape was the source of the *o'* part for the construction of the *oo'*, *o'o*, *no'* and *o'n* distractors. For two of the eight sets of stimuli, choice of *oo* and *nn* items was not determined randomly. Instead, items were chosen as the three stimuli and as the *nn* item tested with a particular *oo* item on the basis of the lowest possible confusions made between any items seen together in the stimulus list or the test set.

The 24 stimuli and 16 test sets (with positions of items randomly determined in each set) were photographed and reduced to 35 mm negative slides. A set of three new practice shapes and two test sets, made in the same way as other test sets, were produced and shown as 35 mm negative slides. Shapes were more similar to the test shapes to give subjects more realistic practice with the items. Each shape consisted of a symmetrical figure of four line segments enclosed within a bi-axially symmetrical 16-sided polygon. The practice and test stimuli are presented in Appendix 5. A stimulus list and its two test sets was shown in Figure 11.

Apparatus

Stimuli were back projected onto a large screen directly in front of the subject by a Kodak Carousel slide projector, Model S-AV 2000, installed with an electronic timing device to display a fixed number of slides for a specified presentation time and ISI. Another Kodak Carousel slide projector, Model S-RA, was placed next to the first to project the test sets on the screen, slightly to the right of the stimuli. The visual angle for subjects was 14°, with the subject placed approximately one metre from the centre of the screen.

Procedure

The procedure was the same as that of Experiment 4, except that subjects were given half the stimulus lists (four sets of three shapes) at one presentation rate and the other half at the other presentation rate in a counterbalanced order. First, the three practice items were shown to subjects at the same presentation rate as that which was to be used for the presentation of their first four lists. Then the two practice test sets were shown successively, following the same procedure described in Experiment 4.

The first four lists were presented at one presentation rate and tested. Then the subjects were shown the three practice pictures again at the second rate in order to prepare subjects for the change in presentation times. There was no memory set shown after this second presentation of the practice pictures. Then the last four sets of pictures were shown and tested. The order of the four sets of stimuli was randomized for each subject within each presentation time condition, as well as the order of the three stimuli within each list and the order of their two test sets. In this experiment, four sets of pictures were assigned to test each presentation condition. The rated confusability of pictures tested within any set was noted and sets were matched across conditions. It was realized afterward that it would have been better to assign all stimulus sets randomly to presentation conditions, and this improvement was instituted in the experiments to follow. After recognition testing, the questionnaire used in Experiment 4 was administered, and subjects were also asked if they noticed whether presentation time affected their performance.

7.4.2 Results

Means and standard deviations of the proportions of responses to the *oo*, *oo'*, *on*, *no* and *nn* items are shown in Table 22 for each presentation condition. Means and standard deviations of CCP and TP scores are also shown for each condition. Proportions of both the hit rate and

Table 22

Means and standard deviations of recognition scores in each experimental group in Experiment 5.

Score		Presentation Time Group	
		2 Seconds	5 Seconds
oo	$\bar{X}$	.4188	.5000
	S.D.	.1476	.1404
oo'	$\bar{X}$	.2563	.2438
	S.D.	.1375	.1249
on	$\bar{X}$	.1563	.1188
	S.D.	.1136	.1179
no	$\bar{X}$	.1250	.1125
	S.D.	.0995	.1208
nn	$\bar{X}$	.0375	.0250
	S.D.	.0714	.0510
TP (*oo* + *oo'*)	$\bar{X}$	.6750	.7438
	S.D.	.1643	.1179
CCP $\left(\frac{oo - oo'}{oo + oo'}\right)$	$\bar{X}$	.2463	.3486
	S.D.	.3578	.2988

and TP scores were higher in this experiment than they were in Experiment 4. The *nn* rates for this experiment were comparable to those in the previous experiment. The estimated guessing rate was 22.5% for the 2 Second presentation time condition and 15% for the 5 Second condition. Subtracting the guessing rates from *oo* scores still showed that subjects remembered at least a quarter of the stimuli accurately under both conditions. CCP scores were much higher in this experiment than in Experiment 4, as a consequence of the higher hit rates. Thus, the fragmented memory effect was not as marked in subjects' performance in this experiment.

Hit Rate Analysis

To determine whether subjects performed more accurately in the 5 Second condition than in the 2 Second condition, hit rates for the two conditions were compared. Subjects chose significantly more *oo* responses under the 5 Second presentation condition than under the 2 Second condition ($t = 1.8185$, $df = 19$, $p < .05$ for a one-tailed test). Thus, subjects remembered more pictures when given longer viewing time.

The order of presentation of the testing conditions did not affect accuracy scores. Hit rates for the 2 Second condition averaged 4.0 for those presented with these items first and 2.8889 for those given this condition last, but this was not a significant difference ($t = 0.5893$, $df = 18$, $p > .05$ for a two-tailed test). Subjects given the 5 Second condition first averaged 3.8889 hits, while subjects tested for this condition last averaged 4.0909 hits. There was not a significant difference between these means ($t = 0.3909$, $df = 18$, $p > .05$ for a two-tailed test).

The Fragmented Memory Effect

To determine whether subjects in the more accurate 5 Second condition were less affected by the fragmented memory effect than they were with the 2 Second condition, CCP scores were scored for the two conditions. There was no significant difference between mean CCP scores in the two conditions ($t = 1.1171$, $df = 19$, $p > .05$ for a two-tailed test). Thus, given that subjects remembered the parts of the stimuli, they were as likely to identify the *OO* items in the test set with two seconds of presentation time as with five seconds.

To determine whether subjects made choices of *OO* items more often than *OO'* items in this experiment, each mean CCP score for each condition was tested to see if it differed significantly from zero. A single-mean t-test yielded a t-value of 3.0788 for the 2 Second condition and 5.2186 for the 5 Second condition. Both values exceed the critical value of t at the 1% level of significance. Hence, in both conditions subjects could, to some extent, discriminate between the correct item and *OO'* distractors.

Sex Differences

No sex differences were found. Males averaged 3.9 hits and females averaged 3.45 hits, but this difference was not significant ($t = 1.2077$, $df = 18$, $p > .05$ for a two-tailed test).

Confidence Ratings

Means and standard deviations of confidence ratings for each item type in both conditions are shown in Table 23. Subjects were more confident of *oo* than *oo'* responses. For all hits, insides were rated with more confidence than outsides. For *on* items, outsides were rated with less confidence than insides, while the reverse trend was observed for *no* items. For both experimental groups, "whole" ratings for *no* items were lower in confidence than ratings for *on* items, replicating a trend found in Experiment 4. As expected, *nn* items were rated with less confidence than other items, especially in the more accurate 5 Second condition. Subjects were, to some extent, aware of the accuracy of their responses.

Questionnaires

Answers to the post-recognition questionnaire gave further support to the trends in the confidence ratings. No subjects reported concentrating on outsides of shapes, while 55% reported concentrating on the insides. Some subjects (15%) said they could only concentrate on one part, without stating a preference, and one more subject stated that he concentrated on only one portion of each shape which included some of the inside as well as the outside part. Other subjects (10%) noted that they were more able to see both parts of the stimuli at the longer presentation time. Most subjects (60%) reported that the longer presentation time was better for remembering than the shorter, while 35% of subjects noticed no change in their feeling of subjective accuracy over the two presentation times. Subjects performed more accurately with longer presentation times, and their confidence ratings for *oo* and *nn* items, as well as their comments, showed that they knew that they were more accurate on the pictures presented for five seconds than those presented for two seconds.

7.4.3 Discussion

The present study, using longer presentation times than Experiment 4, supported findings of other investigators (see section 6.2)

Table 23

Means and standard deviations of confidence ratings for insides (I), outsides (O) and wholes (W) of all response types in each experimental group in Experiment 5.

Group		Response Type														
		oo			*oo'/o'o*			*on/o'n*			*no/no'*			*nn*		
		I	O	W	I	O	W	I	O	W	I	O	W	I	O	W
2 Seconds	$\bar{X}$	1.51	1.90	1.93	1.62	2.17	2.31	1.56	2.40	2.32	2.35	2.00	2.50	2.00	2.00	2.33
	S.D.	0.75	0.82	0.72	0.70	0.79	0.72	0.71	0.58	0.56	0.81	0.80	0.69	0.89	0.89	0.82
	frequency		67			42			25			20			6	
5 Seconds	$\bar{X}$	1.41	1.68	1.71	1.64	2.05	2.10	1.74	2.05	2.11	2.22	2.06	2.50	2.75	3.00	3.00
	S.D.	0.71	0.74	0.73	0.71	0.86	0.75	0.93	0.78	0.74	0.73	0.87	0.52	0.50	0.00	0.00
	frequency		80			39			19			18			4	

which showed that increasing presentation time leads to more accurate recognition. The two presentation time conditions yielded remarkably similar patterns of errors: their proportions of *oo'*, *on*, *no* and *nn* responses are comparable. The longer presentation time may only decrease the probability that any type of error will occur, with slightly more decrease in *on* errors than any others. Since most subjects reported that they attended first to the inside part, then to the outside, the longer presentation time may simply ensure that the outside can be processed during the presentation time.

The findings of the present study, in conjunction with findings from the previous study, suggest that increasing the presentation time from 0.5 to two seconds has no effect, but increasing it from two to five seconds does allow subjects to encode more information to aid their recognition. However, the stimuli and distractors were different in the present experiment from those used in the previous experiment, and this alone may have produced the differences in presentation time effects. Indeed, the previous experiment included a group given two seconds presentation time with two seconds ISI. The performance of this group was not as accurate as that of the group given two seconds presentation time and 1.5 seconds in the present experiment, and it seems as though methodological differences between the two studies could have accounted for the differences in their findings. The last experiment to be described in this chapter (Experiment 7) attempts to replicate the findings of the first experiment with the new stimuli and test sets.

The stimuli in the present experiment were easier to recognize, at least partly because items in each set were as discriminable as possible. The higher CCP scores also indicate that subjects had much less difficulty with these stimuli than subjects in the previous experiment. Although *oo'* responses occurred much more than other errors, they occurred somewhat less often than they had in Experiment 4. The fragmented memory effect may therefore be a function of the confusion between items in the recognition sets, rather than a function of the amount of encoding time. The tendency cannot be attributed to chance responding, since *oo'* scores occurred about 8% more often then would have been expected by chance, while *on* and *no* responses had much lower frequencies which could have been attributed to chance performance.

7.5 EXPERIMENT 6: THE EFFECT OF LENGTHENING ISI ON RECOGNITION OF SHAPES

To determine whether the time between pictures in a list can be used to benefit memory, in the present experiment a short (1.5 second) ISI was compared with a long (seven second) ISI in a recognition task concerned with measuring memory for whole abstract shapes as well as their constituent parts. The presentation time was held constant at two seconds for both the 1.5 Second and 7 Second conditions. This was done so that the 1.5 Second group could be compared with a similar group in Experiment 5, to determine the reliability of the memory estimates found in these experiments. The 7 Second condition was chosen because previous studies reviewed (section 6.2) had not tested ISI's longer than four seconds. The large difference in times given for the ISI's was chosen to maximize any differences in recognition accuracy due to this temporal variable.

The procedure was improved in this experiment. A between-groups design was employed to increase the data collected from subjects in each group. Also, the practice test sets were changed so that *oo'* items were not included. Instead, a new part was used to create a second *no* or *on* distractor in each of the practice test sets. This was done to prevent subjects' guessing during practice that parts from different stimuli would be put together to create foils for the test sets. This might lead to test-wise strategies on the part of subjects to aid their linking the inside and outside parts together in an artificial way.

7.5.1 Method

Subjects

Twenty-eight first-year Psychology students volunteered to be subjects. Fourteen subjects were assigned to each of the two conditions in a counterbalanced order. Two subjects in each condition failed to respond to all the test sets, but only one of these failed to respond more than once and failed to make the appropriate confidence ratings (i.e., he did not follow instructions and failed to give any response to several test sets) and was rejected, leaving only 13 subjects in the 7 Second condition for analysis.

Procedure

Stimuli and apparatus were the same as those used in Experiment 5. Instructions were also the same, except that subjects were told that all pictures would be presented at the same rate. A different practice set was employed in this experiment. During practice, subjects were shown three practice stimuli, followed by their two test sets. In each set, the *oo'* item was replaced by either a *on* or an *no* item with another new part. One set had two *no* items with two different new insides and the other had two *on* items with two different outsides.

After recognition testing, subjects were asked the questions used in the post-recognition interview in Experiment 5.

7.5.2 Results

Means and standard deviations of proportions of responses to *oo*, *oo'*, *on*, *no* and *nn* items and the CCP and TP scores are presented in Table 24 for each ISI group. Results were similar to those found for Experiment 5. The hit rate for both groups was high, as well as the combination of *oo* and *oo'* responses (TP scores), indicating that memory for parts of the stimuli as well as whole pictures was adequate. Subjects in the 7 Second ISI group had somewhat higher *oo* and TP scores than subjects in the 1.5 Second ISI group. Also, their guessing rate was lower, since they were estimated to have been guessing on 15% of occasions, while the 1.5 Second subjects were guessing on about 19% of occasions. CCP scores were more positive for the 7 Second ISI group as a consequence of their higher hit rates.

Hit Rate Analysis

To test whether subjects in the 7 Second group were more accurate in recognition than subjects in the 1.5 Second group, mean *oo*

Table 24

Means and standard deviations of all recognition scores in each experimental group in Experiment 6.

Response		Interstimulus Interval Group	
		1.5 Seconds	7 Seconds
oo	$\bar{X}$	.4471	.5003
	S.D.	.1507	.1196
oo'	$\bar{X}$	.2463	.2474
	S.D.	.1118	.1122
on	$\bar{X}$	.1626	.1266
	S.D.	.1308	.0894
no	$\bar{X}$	.1080	.1010
	S.D.	.0831	.0866
nn	$\bar{X}$	.0316	.1247
	S.D.	.0412	.0424
TP (*oo* + *oo'*)	$\bar{X}$	.6934	.7478
	S.D.	.1857	.1109
CCP $\left(\frac{oo - oo'}{oo + oo'}\right)$	$\bar{X}$	.1407	.3431
	S.D.	.3782	.2498

scores in the two groups were compared. No significant difference was found ($t = 1.0057$, $df = 25$, $p > .05$ for a two-tailed test). Thus, subjects in the two groups were comparable in their hit rates. No advantage of a longer ISI was found.

The Fragmented Memory Effect

To determine whether subjects in the 7 Second group chose correct combinations of parts more often when they remembered both parts of a stimulus, the independent means of CCP scores were compared for the two groups. No significant difference between the groups was found ($t = 1.6257$, $df = 25$, $p > .05$ for a two-tailed test). It seems that subjects in the 7 Second group were no more accurate in choosing correct combinations than subjects given 1.5 seconds ISI. This is consistent with the results of the hit rate analysis.

To determine whether subjects in each group could distinguish between *oo* and *oo'* items, for each group a t-test was carried out on the mean CCP score of each group individually, to see if it differed significantly from zero. The 1.5 Second ISI group did not have a score that differed significantly from zero ($t = 1.1707$, $df = 13$, $p > .05$ for a two-tailed test), while the positive mean CCP score of the 7 Second group was significantly greater than chance ($t = 4.7587$, $df = 12$, $p < .002$ for a two-tailed test). This suggests that subjects in the 7 Second group were able to choose the correct test item when both parts of a picture were remembered, while 1.5 Second subjects could not, and were more influenced by the fragmented memory effect. Since the results of these analyses are contradictory, it is unclear whether the groups actually differed from each other in their tendency to choose correct combinations of parts. Hence any interpretation of these results should be cautious.

Memory for Only One Part of a Picture

To determine whether subjects remembered more of the inside parts than outside parts of each shape, mean proportions of subjects' *on* and *no* responses within each experimental group were compared. No significant difference was found between the proportions of these responses in either the 1.5 Second group ($t = 1.5116$, $df = 25$, $p > .05$ for a two-tailed test) or the 7 Second group ($t = 0.785$, $df = 25$, $p > .05$ for a two-tailed test). Thus, it appears that subjects in both groups remembered as many inside parts as outside parts when they remembered only one part of a stimulus.

Delay

To determine whether forgetting was affected by delay between list presentation and testing, memory scores for the first test sets shown after each list were compared with memory scores for the second sets. Subjects had controlled their own pace in viewing these two sets in succession, so delay was not the same for each subject, although the amount of interference from other materials was constant for both the first sets shown and the second sets shown (though there was more interference for the second set than the first in both cases). Means and standard deviations of proportions of *oo* responses for the first and second test sets shown are presented in Table 25 for each ISI group. Each subject's *oo* scores were used to compare memory performance on the first test sets with performance on the second sets. The order of presentation of test sets did not affect *oo* scores in either the 1.5 Second group ($t = 1.7717$, $df = 13$, $p > .05$ for a two-tailed test) or the 7 Second group ($t = 0.0842$, $df = 12$, $p > .05$ for a two-tailed test). Thus, the combined influence of longer delays and interference from the first test set did not significantly alter memory scores on the second set.

Table 25

Means and standard deviations of *oo* scores for each group on the first and second distractor set: Experiment 6.

oo Scores		Group	
		1.5 Seconds ISI	7 Seconds ISI
First set	$\bar{X}$	0.5928	0.4961
	S.D.	0.1960	0.1679
Second set	$\bar{X}$	0.4072	0.5039
	S.D.	0.1960	0.1679

Confidence Ratings

Confidence ratings for each type of response in the test sets for the two experimental groups are shown in Table 26. These replicate trends in the earlier studies. Notably, the insides of shapes were rated with more confidence than outsides on every item except *no* items.

Questionnaires

Results of the post-recognition questioning of subjects supported previous findings of experiments reported earlier in this chapter. The majority of subjects, 79% of 1.5 Second ISI subjects and 92% of 7 Second ISI subjects, reported that they concentrated on the insides of shapes or attended initially to them. Also, most subjects verbalized during the presentation of the shapes (e.g., "they looked like faces"; "a bear was one"); 71% of subjects in the 1.5 Second group and 76% of

Table 26

Means and standard deviations of confidence ratings for insides (I), outsides (O) and wholes (W) of each test item for each condition in Experiment 6.

ISI Group	Response Type														
	oo			*oo'/o'o*			*on/o'n*			*no/no'*			*nn*		
	I	O	W	I	O	W	I	O	W	I	O	W	I	O	W
1.5 Seconds															
$\bar{X}$	1.90	2.66	2.85	2.19	2.76	3.22	1.92	3.36	3.44	3.38	2.63	3.92	3.14	4.00	4.57
S.D.	1.30	1.38	1.30	1.29	1.21	1.30	1.32	1.15	1.18	1.10	1.28	1.02	1.07	1.53	1.13
frequency		98			54			36			24			7	
7 Seconds															
$\bar{X}$	1.65	2.44	2.58	1.67	2.35	2.69	1.38	3.19	2.85	3.10	2.90	3.71	2.80	3.40	3.80
S.D.	1.03	1.27	1.34	0.99	1.45	1.38	0.98	1.17	1.19	1.22	1.89	1.23	2.05	1.34	1.79
frequency		103			51			26			21			5	

subjects in the 7 Second group verbalized during the presentation of pictures. Several subjects in each group (44% in all) spoke spontaneously of attempting to use a strategy to link the inside and outside parts together in memory. Finally, rehearsal was more often reported in the 7 Second group than in the 1.5 Second group, which would be expected if subjects require a certain amount of time to use a rehearsal strategy for visual material in the interval. Most (69%) of the 7 Second subjects reported rehearsing in the ISI, while only 14% of 1.5 Second subjects reported any rehearsal.

Confusability

Some of the test sets contained shapes rated as more confusable with each other than other sets (see section 7.3.1). To discover whether the more confusable sets had lower recognition scores, a Spearman rank correlation coefficient was calculated between the number of ratings of confusion per set (over all judges) and the mean number of hits (*OO* scores) found for each set in the experiment. An r_s of 0.438 was found, which is significantly above chance at the 5% level. This shows that sets given higher ratings of confusability were more difficult to recognize. This confirms the general finding in studies of recognition memory for pictures that recognition distractors more similar to the test item are more difficult to reject.

7.5.3 Discussion

The results of the present study suggest that recognition performance is no better when seven seconds of ISI are given between successive stimuli than when 1.5 seconds of ISI are given. This suggests that subjects can only use the interval between stimuli to maintain the amount of information which has been encoded during the presentation time. On the other hand, one study [Tversky and Sherman, 1975] has found that recognition performance improves when ISI is lengthened from 1.5 to three seconds. Perhaps increases in ISI do aid memory, but only when they are of short duration. With longer ISI's, the increased delay between presentation of stimuli and recognition testing may lead to forgetting which outweighs any advantages of extra processing time in the ISI. The only evidence concerning delay in the present experiment was the finding that subjects performed as adequately on the second recognition test set as the first, despite the interference from other materials and delay which was greater for the second set. This finding does not support the contention that delay had a negative effect on recognition in the present experiment. This may be support for Cermak's [1971] finding that not much information is lost from memory between four seconds and 20 seconds after a shape has been presented. The initial information lost in the first four seconds may reflect a short-term limit in memory. After the first few seconds, the information available may be transferred to a more stable, long-term store which is less affected by decay. This, however, is only a tentative interpretation of the results of the present experiment. In fact, delay was not found to affect hit rates in the present experiment, either when performance on the first recognition sets was compared with performance on the second sets given, or when performances of groups given variable amounts of time to view each list (i.e., varying ISI's) were compared. Whether the lack of delay effects was due to an active rehearsal strategy on the part of the subjects given seven seconds rather than 1.5 seconds in the ISI or whether it was due to the stability of a passive memory trace cannot be determined from the data.

One criticism of the present experiment is that numbers of subjects in each group were smaller than was desirable, and that a more powerful test might have found a significant advantage of the seven second ISI condition over the 1.5 second condition. However, even when subjects' *OO* scores in the 7 Second ISI group were compared with the larger number of scores from subjects given 1.5 seconds ISI (and the same two seconds for presentation time) in Experiment 5, still no significant difference between the groups was found. This adds support to the hypothesis that increases in ISI do not aid memory processing.

Fragmented memory did have some effect on memory performance in the present experiment. However, subjects in both groups were able to choose the correct item in recognition at least twice as often as distractors made by recombining parts seen in different stimuli (*OO'* items). This is similar to the level observed in Experiment 5. It means that subjects who are shown pictures for at least two seconds each, with ISI's ranging from 1.5 to seven seconds, can successfully discriminate correct whole pictures from incorrect combinations of correct parts on perhaps one-quarter of occasions. (This figure was calculated by subtracting the proportion of *oo'* scores from the proportion of *oo* scores in each group, since when subjects remember only parts of pictures without remembering their combinations, they should choose randomly between *oo* and *oo'* items.) This estimate of memory accuracy is not high. It implies that even with increased encoding time, there is still a strong fragmented memory effect, and indeed, in all experiments so far reported in this chapter, *oo'* distractors have been chosen much more often than any other distractors. This shows that subjects do have information about parts of pictures seen in the recognition stimulus set, even if whole combinations are not perceived or are quickly forgotten.

The finding that *no* items were not appreciably easier to reject than *on* items on the recognition task replicates a finding of Experiment 5, and suggests that the stimuli were not biased in having more interesting or perceptible inside figures. Post-recognition questioning showed that most subjects still reported that they attended to the inside figure most or noticed it first, but the presentation conditions must have been long enough to allow encoding of both parts. There may be a tendency towards looking to the centre of a picture initially. This may have been a wise strategy in the present experiment, since the shapes occupied a wide visual angle and subjects not attending to the centre of the screen may have lost valuable time in scanning the picture. Fixation points were not used for two reasons:

(1) their initial presentation may have interfered with the first picture presented, and

(2) it was desired to study memory for pictures when subjects were allowed to encode the pictures under normal viewing conditions.

Confidence ratings complemented results of the post-recognition questionnaire in this experiment. Subjects were more confident of inside than outside parts of both *oo* and *oo'* items chosen. However, since they were also more confident about insides than outsides in the *nn* items chosen, there appears to have been an unrealistic response bias towards accepting insides of shapes. This also occurred in

Experiment 5, but only for the group given five seconds presentation time. In all groups in Experiments 5 and 6, *on* responses were given more confident ratings as "wholes" than *no* responses. This again suggests that subjects were more confident about their choices of correct insides, even though their choices of correct insides and outsides appeared to be equally accurate.

The reliability of the findings using this recognition task is shown in the comparison of results for the groups given 1.5 seconds ISI and two seconds presentation time in Experiments 5 and 6. All proportions of responses are remarkably similar. CCP scores were lower in Experiment 6 than in Experiment 5, but the frequency of subjects unable to discriminate between *oo* and *oo'* items (with CCP scores of zero or lower) was approximately the same in both studies. The consistency gives support to the reliability of the recognition method used to test differences, despite some methodological differences between studies, and suggests that subjects were able to employ consistent encoding processes during the task.

The significant correlation found between confusability of items in test sets, based on ratings carried out in Experiment 5, and the number of hits subjects in both groups made on these sets is further evidence that it is extremely important to control the similarity of items tested in a recognition task. The difficulty of the recognition task was better controlled by confusability ratings in Experiments 5 and 6 than by the verbalizability and unity-of-parts ratings used in Experiment 4. In the next section, Experiment 4 is replicated with the new stimulus and test sets controlled for the confusability between items.

7.6 EXPERIMENT 7: A REPLICATION OF EXPERIMENT 4

Subjects in Experiment 4 had great difficulty with the recognition task and were easily confused about the stimuli. A change in the method of selecting stimuli and distractors led to increased accuracy in Experiments 5 and 6. The stability of results in the improved method was demonstrated by the close replication of findings for two different groups given two seconds presentation time and 1.5 seconds ISI in Experiments 5 and 6. Experiment 4, on the other hand, stands apart in its findings for a similar group given two seconds for both presentation time and ISI, and for its failure to find the significant effect of presentation time on recognition accuracy which has been consistently demonstrated in previous studies. Experiment 7 was a replication of the fast presentation conditions of Experiment 4 with the improved method.

The present experiment was a replication of the design of Experiment 4, except that the group that had been given two seconds for both presentation time and ISI in Experiment 4 was omitted (because this condition was similar to groups used already in Experiments 5 and 6). Both presentation time and ISI were varied at two levels, with an 0.5 second rate and a two second rate, to test the effects of the length of presentation time, ISI and "total time" on recognition of parts of shapes and their combinations.

The present experiment provided a test of the "total time hypothesis". This hypothesis suggests that presentation time and ISI are equivalent in their effects on memory, and that the amount of time elapsing between onset of one picture and onset of the next picture in the list is the only important temporal factor affecting recognition. This would occur if subjects were able to make as much use of time after stimulus offset in their encoding of information, as during stimulus presentation. Thus, the total time hypothesis suggests that subjects given 0.5 seconds presentation time and two seconds ISI would perform as well on the memory task as subjects given two seconds presentation time and only 0.5 seconds of ISI. Alternatively, if presentation time is the most important factor in recognition, the group given the longer (two seconds) presentation time should perform better than groups given an 0.5 second presentation time regardless of the ISI. If ISI is important in increasing recognition, subjects given two seconds ISI with 0.5 seconds of presentation time should perform better than subjects given 0.5 seconds ISI with the same presentation time.

7.6.1 Method

Subjects and Design

Fifty-one first-year Psychology students volunteered for the experiment. All were paid for their participation. Seventeen subjects were assigned randomly to each of the three experimental groups with one of the following presentation conditions:

(1) 0.5 seconds presentation time with 0.5 seconds ISI (the 0.5/0.5 group),

(2) 0.5 seconds presentation time with two seconds ISI (the 0.5/2 group,

(3) two seconds presentation time with 0.5 seconds ISI (the 2/0.5 group).

Stimuli

The 24 stimuli and 16 test sets were those used in Experiments 5 and 6. Stimuli were presented in lists of three, which were tested by two recognition sets. The order of lists presented, the order of stimuli in each list and the order of the test sets shown after each list were randomly determined for each subject.

Apparatus

For fast, successive presentation of the three pictures in each stimulus list, three Kodak Carousel slide projectors, one Model S-AV 2000 and two Model S-RA, were arranged side by side so that they projected slides onto the same area of a back projection screen. The visual angle for subjects was the same in this experiment as in Experiments 5 and 6 (14°). A modular, digital electronic timing device was connected to three solenoid shutters so that each shutter opened for a specified amount of time in succession, and the length of time between closing of one shutter and the opening of the next in the sequence of three could also be controlled precisely. Each shutter was mounted in front of the lens of one of the projectors so that exact presentation time and ISI could be controlled for each list of stimuli. One of the

Model S-RA projectors was used to present the slides of the test sets after each list of stimuli was shown.

Procedure

The procedure for subjects was the same as that of Experiment 6. Subjects were tested individually, seated before the rear projection screen.

7.6.2 Results

Means and standard deviations of proportions of responses to each type of item in the test sets, TP scores and CCP scores are shown in Table 27. The hit rates, TP scores and CCP scores all appear to have increased with both longer presentation times and ISI's, though the longer presentation time seems to have improved recognition accuracy more than the longer ISI. The *oo'* rate was stable across conditions in the experiment, as it had been in all the earlier experiments reported in this chapter. The frequency of choices of distractor items containing new parts decreased with longer hit rates found. Evidence that recognition accuracy increased with longer total time was provided by differences in *nn* responding. It was estimated that the guessing rate was 45% in the 0.5/0.5 group, while it was less than 18% in both the other groups which were given longer total time to process each stimulus.

Hit Rate Analysis

To determine whether there were differences between groups in hit rate, a one-way analysis of variance (random effects) was carried out on the *oo* scores for subjects in each group. The summary table is shown in Table 28. There was a significant effect of presentation condition on hit rates ($p < .01$). A Newman-Keuls test for *post hoc* comparisons carried out on the mean scores for each group found that subjects in the 2/0.5 group had significantly higher *oo* scores than subjects in the 0.5/2 group ($p < .01$), and that these latter subjects had significantly higher *oo* rates than subjects in the 0.5/0.5 group ($p < .05$).

Table 27

Means and standard deviations of all response scores for each experimental group in Experiment 7.

Score		Presentation Condition		
		0.5/0.5	0.5/2	2/0.5
oo	$\bar{X}$	.2536	.3104	.4352
	S.D.	.1122	.0714	.1706
oo'	$\bar{X}$	.2610	.2731	.2747
	S.D.	.1068	.0922	.1122
on	$\bar{X}$	.2145	.2169	.1552
	S.D.	.1058	.1136	.1095
no	$\bar{X}$	.1995	.1738	.1079
	S.D.	.1082	.1414	.0729
nn	$\bar{X}$	.0756	.0257	.0270
	S.D.	.0640	.0387	.0400
TP (*oo* + *oo'*)	$\bar{X}$	.5146	.5791	.7100
	S.D.	.1249	.0990	.1400
CCP $\left(\frac{oo - oo'}{oo + oo'}\right)$	$\bar{X}$	-.0304	.0760	.2071
	S.D.	.3633	.2309	.3363

Table 28

Summary of the analysis of variance for hit rates: Experiment 7.

Source	SS	df	MS	F	p
Presentation Condition	0.2933	2	0.1467	9.4038	< .01
Error (within groups)	0.7484	48	0.0156		
Total	1.0417	50			

Thus, extending the presentation time from 0.5 to two seconds had significantly more effect on memory accuracy than extending the ISI over the same range. This refutes the total time hypothesis and suggests that, though increases in either presentation time or ISI can lead to greater accuracy of memory, presentation time is more important in processing stimuli for memory than the ISI.

The Fragmented Memory Effect

To determine whether subjects who remembered parts of pictures correctly could recognize their combinations, CCP scores were subjected to the same analysis as the *oo* scores. A one-way analysis of variance was carried out on the mean CCP scores for subjects in each group and is summarized in Table 29. There was no significant effect of presentation condition. Thus, groups did not differ in their tendency to make confusions between pictures. To determine whether CCP scores differed significantly from zero, a t-test was carried out on the mean CCP score for each group. The average CCP score did not differ significantly from zero in either the 0.5/0.5 group ($t = 0.3348$, $df = 16$, $p > .05$ for a two-tailed test) or the 0.5/2 group ($t = 1.3172$, $df = 16$, $p > .05$ for a two-tailed test). The 2/0.5 group did have a mean CCP score which differed significantly from chance ($t = 2.4625$, $df = 16$, $p < .05$ for a two-tailed test). Since the results of these two analyses are contradictory, they must be treated cautiously. The results of the second analysis suggest that groups given 0.5 seconds of presentation time could not distinguish between *oo* and *oo'* items, while the group given two seconds presentation time had no difficulty in choosing *oo* items over *oo'* items. This can be taken as only a tentative interpretation of the data.

Table 29

Summary of the analysis of variance for
CCP scores: Experiment 7.

Source	SS	df	MS	F	p
Presentation Condition	0.4811	2	0.2406	2.4205	NS
Error (within groups)	4.7731	48	0.0994		
Total	5.2542	50			

The results of this analysis of CCP scores cannot be used to make conclusive statements about the fragmented memory effect, since none of the groups differed significantly from any of the others. However, since the experimental procedure was equivalent for subjects in Experiments 6 and 7, comparisons of mean CCP scores were carried out between the 0.5/0.5 group in the present experiment and the group given

1.5 seconds ISI and two seconds presentation time in Experiment 6 (the 2/1.5 group). The former group was also compared with the group given seven seconds ISI and two seconds presentation time (the 2/7 group) in Experiment 6. Mean CCP scores were significantly higher in the 2/1.5 group than the 0.5/0.5 group ($t = 2.9566$, $df = 29$, $p < .02$ for a two-tailed test). They were also higher in the 2/7 group than the 0.5/0.5 group ($t = 3.069$, $df = 28$, $p < .01$ for a two-tailed test). This shows that subjects given more total time to process each stimulus were more able to choose correct combinations of parts from memory, provided that they remembered the parts. Subjects given fast presentations of pictures could not distinguish between the original stimuli (*oo* items) and distractors made by recombining parts of different stimuli (*oo'* items).

Memory for Only One Part of a Shape

To determine whether *on* scores occurred significantly more often than *no* scores in this experiment, matched pairs of mean *on* and *no* scores for each subject were compared. No significant difference between *on* and *no* responses was found for the 0.5/0.5 group ($t = 0.3264$, $df = 16$, $p > .05$ for a two-tailed test), the 0.5/2 group ($t = 1.1627$, $df = 16$, $p > .05$ for a two-tailed test) or the 2/0.5 group ($t = 1.359$, $df = 16$, $p > .05$ for a two-tailed test). Thus, *on* errors were made as often as *no* errors in all groups. This replicates findings of Experiments 5 and 6, but not Experiment 4, showing that subjects could remember outside parts as well as inside parts of shapes, despite the fast presentation times used.

Delay

To determine whether interference effects and longer elapsed time between presentation and testing would affect recognition performance on the second test set more than on the first set, matched scores for *oo* rates in the two sets were compared in t-tests for each experimental group. Means and standard deviations of *oo* scores on the two sets for each experimental group are shown in Table 30. The *oo* scores for the 0.5/0.5 group were significantly higher for the first set than the second ($t = 2.3479$, $df = 16$, $p < .05$ for a two-tailed test), but not

Table 30

Means and standard deviations of proportions of *oo* and *oo'* scores for first and second test sets in each experimental group in Experiment 7.

Response Type		Group		
		0.5/0.5	0.5/2	2/0.5
oo Scores:				
First Set	$\bar{X}$	.6510	.5529	.4847
	S.D.	.2651	.1603	.1833
Second Set	$\bar{X}$	.3490	.4471	.5153
	S.D.	.2651	.1603	.1833
oo' Scores:				
First Set	$\bar{X}$	.3853	.4382	.5850
	S.D.	.2017	.2571	.2369
Second Set	$\bar{X}$	.6147	.5618	.4150
	S.D.	.2017	.2571	.2369

for the 0.5/2 group ($t = 1.3605$, $df = 16$, $p > .05$ for a two-tailed test) or the 2/0.5 group ($t = -0.3445$, $df = 16$, $p > .05$ for a two-tailed test). Lengthening presentation time or ISI reduced the difference between the first and second test sets. To determine whether the effect of order of the test sets on *oo* scores had a corresponding effect on *oo'* scores, mean *oo'* scores were compared between the first and second test sets for each group. (Means and standard deviations of these scores are shown in Table 30.) The 0.5/0.5 group was also the only group to have significantly more *oo'* scores on the second test set than the first ($t = -2.3453$, $df = 16$, $p < .05$ for a two-tailed test). There were no significant differences between sets for either the 0.5/2 group ($t = -0.9902$, $df = 16$, $p > .05$ for a two-tailed test) or the 2/0.5 group ($t = 1.4796$, $df = 16$, $p > .05$ for a two-tailed test). The group given only 0.5 seconds for both presentation time and ISI was more affected by the order of presentation of the test sets than groups given longer (two second) presentation times or ISI's. Presumably, decay of memory was more rapid for subjects who had not had time to process each stimulus adequately. Results also suggest that fragmented memory had a greater effect on recognition performance after delay and interference from the first test set.

Confidence Ratings

Confidence ratings for parts of shapes and their combinations for each item in the test sets are shown for each group in Table 31. The *oo* items were given slightly more confident ratings than *oo'* items with regard to their combinations ("wholes"). The *nn* items were given the least confident "whole" ratings in the 0.5/0.5 group and the 2/0.5 group, but the 0.5/2 group gave "whole" ratings of lower confidence to *oo'*, *on* and *no* items than to the *nn* items. This indicates that subjects in the 0.5/2 group were uncertain about their choice of recognition items. There was a tendency to rate insides more confidently than outsides in the *oo* and *oo'* items for all groups, and in the *nn* items for the 2/0.5 group. The bias towards acceptance of insides rather than outsides of shapes in the confidence ratings was not as extreme in this experiment as it had been in the previous experiments reported in this chapter.

Questionnaires

Post-recognition questioning of subjects revealed that most subjects concentrated on insides of the shapes, giving further support to evidence of bias in favour of inside parts found in the confidence ratings. A tendency to concentrate on inside parts was reported by 59% of 0.5/0.5 subjects, 76% of 0.5/2 subjects and 71% of 2/0.5 subjects.

A verbal encoding strategy was used by 82% of subjects in the 0.5/0.5 group, 53% of subjects in the 0.5/2 group, and 70% of subjects in the 2/0.5 group. On the other hand, only 12% of subjects in the 0.5/0.5 group used an encoding strategy which attempted to link the inside and outside parts together, while 29% of subjects in each of the other two groups reported this strategy. Rehearsal in the ISI was reported by 47% of subjects in the 0.5/0.5 group, while 58% of subjects in the 0.5/2 group and 70% of subjects in the 2/0.5 group reported rehearsal. This finding is surprising, since it might be expected that subjects given longer ISI's would rehearse more often than subjects given brief ISI's, but adequate rehearsal may depend on sufficient time in which to encode each picture while it is physically present.

Table 31

Means, frequencies and standard deviations of confidence ratings for insides (I), outsides (O) and wholes (W) of each response type for each experimental group in Experiment 7.

Group		Response Type														
		oo			*oo'/o'o*			*on/o'n*			*no/no'*			*nn*		
		I	O	W	I	O	W	I	O	W	I	O	W	I	O	W
0.5/0.5	$\bar{X}$	2.59	2.97	3.14	2.58	3.00	3.16	2.68	3.37	3.42	2.96	2.76	3.31	3.35	3.35	3.80
	S.D.	1.30	1.19	1.16	1.19	1.14	1.18	1.35	1.25	1.13	1.06	1.21	1.03	1.04	1.31	1.11
	frequency		66			69			57			51			20	
0.5/2	$\bar{X}$	2.02	2.83	2.81	2.34	2.89	2.99	2.10	3.03	3.02	2.98	2.87	3.23	2.71	2.29	2.86
	S.D.	1.10	1.22	1.14	1.20	1.07	1.09	1.36	1.14	1.21	1.17	1.12	1.27	0.95	0.95	1.22
	frequency		83			73			58			47			7	
2/0.5	$\bar{X}$	1.75	2.34	2.28	1.99	2.82	2.79	2.15	3.43	3.20	3.07	2.79	3.25	2.57	3.43	3.43
	S.D.	1.14	1.24	1.16	0.99	1.10	0.88	1.23	1.15	1.04	1.33	1.37	1.32	1.51	0.79	1.13
	frequency		111			71			40			28			7	

7.6.3 Discussion

The results of the present study suggest that both presentation time and ISI are important in recognition. Increasing either presentation time or ISI from 0.5 to two seconds leads to better accuracy of memory for whole shapes. However, it is not merely the total time elapsed which is significant, since among subjects given 2.5 seconds total time for each picture in a list, those given two seconds presentation time perform better than those given two seconds ISI. These results fail to replicate findings of Experiment 4. However, data for the 0.5/0.5 group are very similar in the two experiments, even though patterns of responses are quite different in the other groups tested. The differences in results could be due to the different stimulus sets used, or even to the change from 16 mm films to 35 mm slides. However, the greater confusability of items in Experiment 4 is probably the most significant factor in producing the differences observed. The moderately high (0.44) correlation found between confusions in the sets, as determined by raters (Experiment 5), and the hit rates for each set in Experiment 6 suggest that accuracy is low when sets are very confusable. Scores in Experiment 4 may have been too low in accuracy to portray real effects of temporal variables on recognition.

The *nn* scores in this experiment were comparable to those in Experiments 5 and 6, except for subjects in the 0.5/0.5 group, who may have been guessing on 45% of occasions. The difference in accuracy between this group and the other two groups in the present experiment may have been greater than the hit rate scores suggested. This is further evidence that increasing presentation time or ISI leads to increased accuracy in memory.

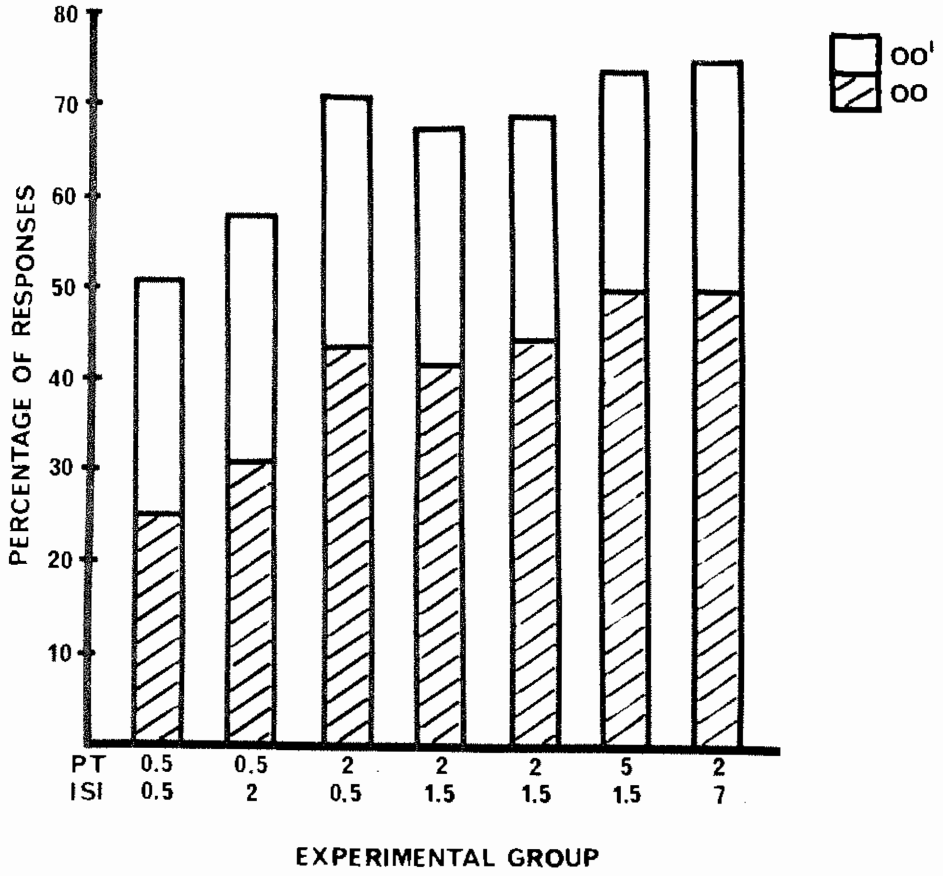
It is interesting that subjects in the 0.5/0.5 group in the present experiment did not choose more *on* than *no* items, as did subjects in a similar group in Experiment 4. If the present subjects were able to see only one part of each picture, then attention was divided equally between the inside and outside parts during encoding. Despite subjects' reports that they were biased towards attending to the insides, memory for the parts did not reflect this bias. This finding opposes a conclusion from Experiment 4 that subjects in the 0.5/0.5 group could only attend to one part of the picture, which, because of subjects' biases, was the inside part. It is odd that, given the same apparent bias, subjects in the 0.5/0.5 group in the present experiment did not remember inside parts better than outside parts. It is not clear whether subjects do attend more to the inside than outside parts. If the outside part can be remembered more easily, subjects may not have to attend directly to it. Since inside and outside parts were different types of configurations in these studies, it is not clear whether the same results would have been found with shapes composed of two similar parts.

In the present experiment, groups did not differ in their CCP scores, though the 2/0.5 group was the only group which had CCP scores significantly higher than chance. There is the tentative suggestion that these subjects could choose the *OO* items more often than the *OO'* items in a list, while subjects given 0.5 seconds presentation time could not differentiate between *OO* and *OO'* items. ISI alone may not affect this tendency, suggesting that the extraction of information about the combination of parts in a shape is carried out while the picture is present. Having two seconds ISI is no more beneficial than 0.5 seconds ISI if the combination of parts has not been encoded during the short presentation time. The 0.5/0.5 group had significantly more *OO* responses on the first test sets than on second sets and significantly more *OO'* responses on second test sets than on first sets. This suggests that the delay and/or interference from other memory tests increases the tendency to encode parts of events without their context. This could mean that event encoding of the whole had taken place, but was quickly forgotten in the 0.5/0.5 group. Groups with longer presentation times and ISI's had more stable memories less influenced by delay and interference.

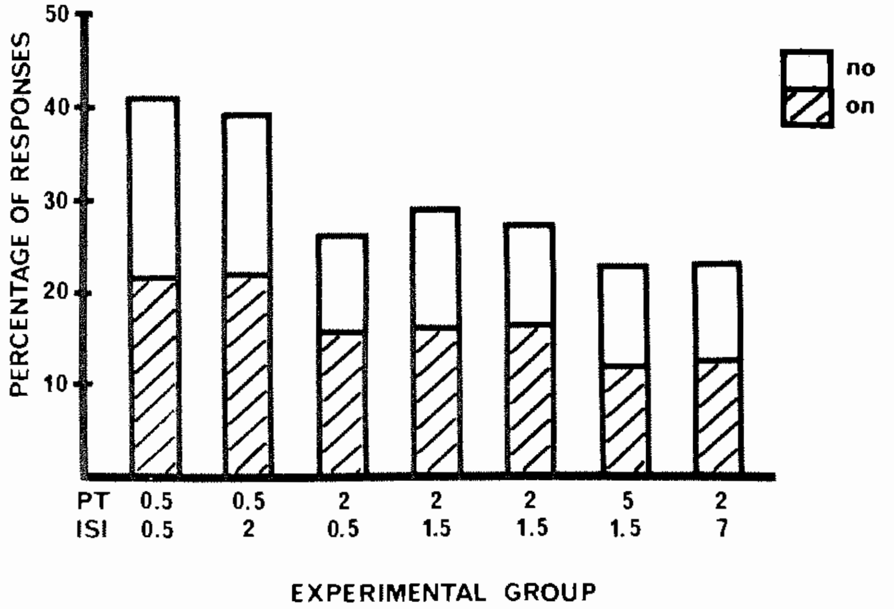
7.7 GENERAL DISCUSSION

7.7.1 Memory for Whole Pictures

These experiments extend previous findings concerning effects of temporal parameters on recognition accuracy. A summary of Experiments 5, 6 and 7 is presented in Figure 12. Mean proportions of



A. Responses with two correct parts



B. Responses with one incorrect part

Figure 12. Percentages of: (A) both *oo* and *oo'* (including *o'o*) responses, and (B) both *on* (including *o'n*) and *no* (including *no'*) responses, for all experimental groups in Experiments 5, 6 and 7. Two groups were given 2 seconds presentation time (PT) and 1.5 seconds interstimulus interval (ISI): the column on the left represents a group in Experiment 5 and the column on the right represents a group in Experiment 6.

responses to *oo*, *oo'*, *on* and *no* items are shown for all experimental groups. It can be seen that the recognition hit rate for two-part shapes increases as presentation time increases from 0.5 to two seconds, and from two to five seconds. This confirms the general finding from research reviewed earlier [e.g., Lutz and Scheirer, 1974; Mooney, 1960] that the hit rate increases as presentation time increases from 60 milliseconds to five seconds. The optimum presentation time is likely to be different for different stimuli, and will depend on other factors such as list length and delay before testing. During the presentation of a picture, the subjects begins to encode as much detail as possible and attempts to integrate details into a coherent memory for the event. With longer time available for processing, more information can be extracted and put into a retrievable form.

Once the stimulus has been removed, processing may still continue. The present experiments showed that increasing ISI from 0.5 to two seconds aids recognition, although extending this interval up to seven seconds does not lead to significantly more accuracy of memory. This seems consistent with a study by Tversky and Sherman [1975] which found that ISI aids memory when it is increased from 1.5 to three seconds. Since Tversky and Sherman used dictionary-style drawings as stimuli, the beneficial effects of increasing ISI from 0.5 to two or three seconds may be a general finding rather than one restricted by the narrow range of stimuli used in the experiments reported in this chapter. In this case, results of a recognition task using controlled stimulus shapes complement results of a more traditional recognition study using representational pictures which are more difficult to control for similarity, complexity and codability. The Tversky and Sherman findings were published after Experiment 6 was carried out, and it was unfortunate that the same temporal variables were not used in both experiments. However, since the methodologies of recognition testing were totally different in the two experiments, there would be no *a priori* basis for studying the same temporal parameters. Accuracy can be influenced by many other factors, as was discussed in Chapter Two. This makes comparisons of different experiments difficult. The important point made by the present research is that lengthening ISI can increase recognition accuracy, and that this effect is not related to the longer total time given for each stimulus, thus extending the findings of Tversky and Sherman's [1975] study.

It may be that longer ISI's aid recognition of shapes only up to a point when the benefits of longer processing time are countered by the effects of increasing delay. Memory for some types of pictures may be affected more by delay than others, creating different critical time periods for encoding different types of pictures. Delay is a complex variable which interacts with other variables. In Experiment 7, there was evidence that delayed recognition testing affected only subjects given inadequate total time for each stimulus (both presentation times and ISI's of 0.5 seconds). These subjects performed much better on the first test set given to test each list than on the second. Other groups given a longer presentation time or ISI had equivalent hit rates on the first and second sets. The combined effects of increased delay before testing and interference from the first test set did not hinder the performances of these latter subjects on the second test set. Thus, it was not merely the type of stimuli given which determined the efficacy of lengthened ISI; the interaction of ISI with presentation time seems to have been a further determining factor.

Evidence from Experiment 7 provides information about the interaction of presentation time and ISI. Increasing either presentation time or ISI from 0.5 to two seconds aids recognition. But, given 2.5 seconds time in total to process the stimulus, recognition is more influenced by presentation time than ISI. It seems reasonable to expect that processing during picture presentation is most extensive while the picture is physically present, since memory failure will not affect encoding during presentation time as it will in the ISI.

The present experiments provide evidence that some sort of encoding of a constructive kind is proceeding during the ISI, with only an icon and memory of the stimulus as input. An alternative interpretation may be that subjects are stressed by such short ISI's or have masking interference which disturbs the presentation of the next picture. Since among subjects given two seconds presentation time, those given an 0.5 seconds ISI (Experiment 7) have a hit rate equivalent to that of subjects given 1.5 seconds of ISI (Experiments 5 and 6), this seems an unlikely conclusion. The increase in memory when ISI is varied from 0.5 to two seconds is even more convincing when the presentation time is only 0.5 seconds (Experiment 7). Since the group given 0.5 seconds ISI is more affected by delay between presentation of the first and second test sets than the group given two seconds ISI, it must be concluded that one negative effect of a longer ISI is to increase the time before testing. Despite the fact that subjects given 0.5 seconds presentation time and two seconds ISI have a longer delay before recognition testing, they perform better than the group given 0.5 seconds for both presentation time and ISI. This finding suggests that it is important to distinguish between the ISI and the delay between stimulus presentation and testing. Both variables are delay variables, and will eventually lead to forgetting if they are extended in time. But lengthening ISI may have beneficial effects for encoding delay, especially when it is not used for rehearsal or is filled with interference [cf., Cohen and Granstrom, 1968] is almost always negative in its effect.

The present studies taken together support the tentative conclusions of the review in Chapter Two. Lengthening presentation time does aid recognition, and the present experiments demonstrated that lengthening presentation time from 0.5 to two seconds, or from two to five seconds, leads to higher recognition accuracy. Lengthening ISI also leads to increased accuracy when it is extended from 0.5 to two seconds, but not when extended from 1.5 to seven seconds. This supports Tversky and Sherman's [1975] finding that both presentation time and ISI affect recognition of pictures. Lutz and Scheirer's [1974] finding that ISI has no effect on recognition may have been due to their use of a mask in the ISI, since they failed to find significant differences when varying some of the same temporal parameters varied in the present research.

7.7.2 Partial Memory

There was a clear demonstration of the fragmented memory effect in the experiments reported in this chapter. Only subjects given

0.5 seconds presentation time were totally unable to distinguish between original stimuli and distractors in the recognition task made by recombining parts of original stimuli. On the other hand, subjects given two seconds or more of presentation time, with ISI's of 0.5 or seven seconds, were more often able to make this distinction. Apparently, if the stimulus is not adequately encoded while it is physically present, parts of the stimulus may not be encoded into the context of the entire picture as an event. The time available after the stimulus has been removed apparently cannot be used to encode the parts as a whole in memory. Of course, it may be that with only 0.5 seconds of presentation time, the parts are not *perceived* as an integrated event. But then this implies that sequentially perceived fragments of a stimulus are not tied together unless the subject has time to attend to their relationships and thus perceive the pictorial event as a whole. This possibility was discussed earlier.

Memory performance even in the most accurate groups was not very high, and all subjects appeared to have been influenced by the fragmented memory effect. Since a quarter of the responses in nearly every group were choices of distractors containing remembered parts in an incorrect combination (*OO'* items), at least a quarter of the correct (*OO*) responses may also have been attributable to the fragmented memory effect. When subjects remember only parts of stimuli seen without remembering their correct combinations, they should choose equally between these (*OO* and *OO'*) items. Thus, though subjects given 0.5 seconds presentation time appeared unable to remember whole stimuli accurately, subjects given longer presentation times also failed to encode wholes on a large proportion of occasions. Since this effect occurred even with a presentation time of five seconds, it is doubtful that subjects were simply unable to perceive the shapes as integrated pictures. The shapes appear to have been easier to remember as separate parts than as wholes.

All the experiments indicated that subjects were biased towards attending to the insides of shapes. Though this led to a greater memory for insides of shapes than outsides in Experiment 4, there were no significant differences in memory for parts of shapes in Experiments 5, 6 and 7. Apparently, this was due to the greater discriminability of items in recognition test sets, which made the task easier for subjects in general, and did not prejudice recognition responses in favour of responses to inside parts in the later experiments. It is interesting to note that, though subjects in Experiments 5, 6 and 7 appeared to be able to encode both parts of the stimuli equally well, they said that they attended first to the insides. With a brief presentation time, it seems unlikely that subjects could have remembered both parts of shapes equally well if one part were attended to first, unless acts of attention are very short, with no more than 0.25 seconds given to each part of a shape. Another explanation may be that subjects normally concentrate their attention on the figure in a figure-ground stimulus [cf., Franken and Davis, 1975], but that the ground is encoded adequately without much conscious effort. It is rather odd that the greater memorability of pictures has been attributed to their greater complexity and richness of information in comparison with verbal and other stimuli, when it might also be suggested that this would increase their difficulty. For example, Nelson *et al.* [1974] found that it was no more difficult to remember complex photographs than unembellished line drawings of the same scene. Of course, spuriously inflated accuracy scores due to partial recognition strategies may be more common in very complex pictures. Nevertheless, the extra detail portrayed in pictures (especially in their "background") does not seem to hinder their memorability, but rather to enhance it. Perhaps complex pictures provide a more elaborated form of memory encoding, with more opportunity to retrieve information later through various, easily accessible cues. Since the outside shapes may have been easier to acquire in memory than inside shapes (though few subjects thought that this was the case), there was no way to separate effects of central versus peripheral attentional strategies from effects of different types of stimuli in the present experiments. An experiment in the next chapter examines memory for insides and outsides of shapes which have parts drawn from the same pool of items to examine this more carefully.

7.7.3 Conclusion

The experiments in this chapter demonstrated that recognition of abstract shapes improves in accuracy when viewing time for each picture is increased from 0.5 to five seconds or when the interval between pictures in a list is increased from 0.5 to two seconds. This suggests that subjects can extract more information from a picture when they are given more time to examine it, and that encoding of information about the picture into memory may continue for a brief period after the stimulus has been removed.

The present experiments also provided evidence that the fragmented memory effect is an important source of error in the recognition task studied and that it may be influenced by temporal variables. Subjects given inadequate time to encode each picture are able to remember parts of the stimuli they have seen, but they are unable to remember how the parts were combined in the originals. Subjects given two seconds or more of presentation time are not as prone to confusions between pictures in memory. Adequate encoding of pictures requires a certain amount of processing time.

CHAPTER EIGHT

EFFECTS OF PERCEPTUAL STRATEGY ON RECOGNITION OF SHAPES

8.1 A CHANGE IN STIMULI TO EXAMINE THE FRAGMENTED MEMORY EFFECT AND THE GENERALITY OF FINDINGS OF CHAPTER SEVEN

Attention to global aspects of a scene may be essential for adequate encoding of the event. An initial global interpretation of a scene may be a necessary initial step in processing rather than a final conclusion about parts to which the subjects has attended [cf., Neisser, 1967]. The attention to individual details in the pictures may be guided by this "general impression" of the picture in which only the most outstanding features of the stimulus are noticed.

There was some doubt as to whether the testing method employed in Experiments 4 through 7 allowed subjects to attend to global aspects of each picture in order to encode each shape as an "event". Though the shapes were unified, symmetrical patterns, their "figure-ground" characteristics may have encouraged subjects to perceive them as "figures" only. Since subjects in all the experiments reported a bias in attending to inside parts of the shapes, it may be that subjects attend to the figure in a figure-ground shape at the expense of the ground. Evidence for this is provided by a study of ratings of pictorial characteristics and their influence on memory. Franken and Davis [1975] found that rating a photograph on the extent to which it contained an obvious figure which stood out against the background was a better predictor of errors in recognition of the pictures than ratings of complexity, "pleasingness" or "interestingness". Though subjects in Experiments 5, 6 and 7 may have been biased towards attending to figures rather than to ground, there was no evidence that inside parts were remembered any better than outside parts. In fact, subjects merely seem to have been more uncertain about their memories for outside parts, though their accuracy scores do not reflect this difference in confidence. If subjects did attend less to outside parts than inside parts, and still remembered both parts equally well, it may be that the outside parts used were intrinsically easier to remember than the inside parts.

The purpose of the experiment described in this chapter was to test the generality of findings in Chapter Seven by testing new sets of stimuli in which inside and outside parts of figure-ground shapes were equal in memorability. Since both parts of each shape were drawn from the same pool of 24-sided, bi-axially symmetrical polygons, the resulting nested shapes contained parts equal in complexity which differed only in size. In this experiment, it was hypothesized that if subjects attend to the insides more often than the outsides, then more insides than outsides should be remembered when subjects are given fast presentation times. In this case, the superior memory for inside parts would be due to subjects' typical attentional strategies rather than to the differences in types of shapes presented as figure or ground.

A second purpose of the experiment to follow was to attempt to manipulate subjects' attention to global features of pictures during presentation by varying their attentional strategies with the use of different perceptual orienting instructions. It was expected that encoding of pictures would be more adequate, and fragmented memory less common, when subjects are forced to attend to global aspects of pictures rather than to parts of a figure-ground configuration. One group of subjects was instructed to perceive the two-part, nested shapes in the most usual way, as figures on a ground. Another group of subjects was instructed to perceive the shapes as "doughnut-like" forms lying on a background which was visible through a "hole" in the centre. Thus, subjects in the "Doughnut" group were encouraged to perceive both polygons simultaneously as boundaries of a single surface. In this case, only a single figure-ground relationship is perceived, with the area between the polygons as the figure, and the other areas of the pictorial surface as the background. On the other hand, subjects in the "Figure-Ground" group were encouraged to perceive three different surfaces in each picture: the central figure, the underlying larger figure and the background for the entire shape.

It was hypothesized that the instructions for the Doughnut group would aid encoding of the shapes as wholes, despite the fact that the shapes are not usually perceived in this way and the perceptual strategy might be difficult to acquire. Doughnut instructions should force subjects to encode portions of both the inside and outside polygons at the same time. If only part of the shape could be encoded in one act of attention, then a small cross-section of both parts should be encoded rather than a single part. This should increase the probability that subjects will remember enough distinctive information about the inside and outside parts to increase their ability to choose correct combinations of parts, relative to subjects given Figure-Ground instructions. The findings reported in Chapter Seven suggested that, at fast (0.5 second) presentation times, subjects cannot discriminate between correct and incorrect combinations of stimulus parts in recognition (i.e., between *OO* and *OO'* items). It was hypothesized that this would still hold true for Figure-Ground subjects here, but that Doughnut subjects would be able to choose *OO* items relatively more often than *OO'* items. Fast presentation times of 0.5 seconds were used in the present experiment with equivalent ISI's in order to test the hypothesis. It was predicted that subjects instructed with the Doughnut strategy would be less affected by fragmented memory than Figure-Ground subjects. The CCP score, indicating the tendency to remember parts but not wholes, was used to give a measure of adequate encoding of whole pictures in order to compare the instructional groups. It was hypothesized that the Doughnut group would have higher CCP scores than the Figure-Ground group.

8.2 EXPERIMENT 8: EFFECTS OF PERCEPTUAL STRATEGY ON RECOGNITION OF PARTS AND WHOLES OF SHAPES

8.2.1 Method

Subjects

Twenty-six first-year Psychology students were paid for their participation in the experiment. Thirteen subjects were assigned randomly to each of the two experimental groups.

Stimuli

The pool of outside shapes used in Chapter Seven was expanded so that there were 84 different, 24-sided, bi-axially symmetrical polygons. The shapes were put together into 42 pairs, each of which contained two distinctively different shapes. One shape in each pair was drawn smaller than the other, so that shapes with 8 cm diameters could be drawn inside the shapes with 15 cm diameters. Thus, the resulting pictures were nested shapes, with each part drawn from the same pool of items. According to raters and pilot subjects, the more compact nature of the inside shapes made them look like quite a different set of shapes.

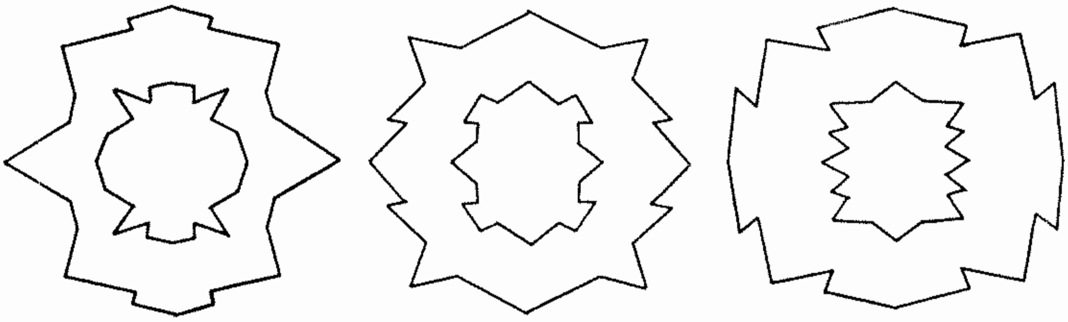
Ten naïve judges were asked to sort the 42 two-part shapes into piles of items which could easily be confused with each other. On the basis of this, eight sets of pictures were made with five least-confusable pictures in each set. More formal estimates of similarity or confusability can be made, but the present informal method (described more fully in section 7.4.1) seemed appropriate for the task.

In each set of five pictures, the three least confusable were designated as stimulus items. The remaining two were designated as *nn* distractors, and test sets were made in the way described in Chapter Seven. One of the stimulus lists and its two test sets are presented in Figure 13. The *oo* and *nn* items least confusable with each other were paired in forming test sets. In cases of equal confusability, items were assigned randomly.

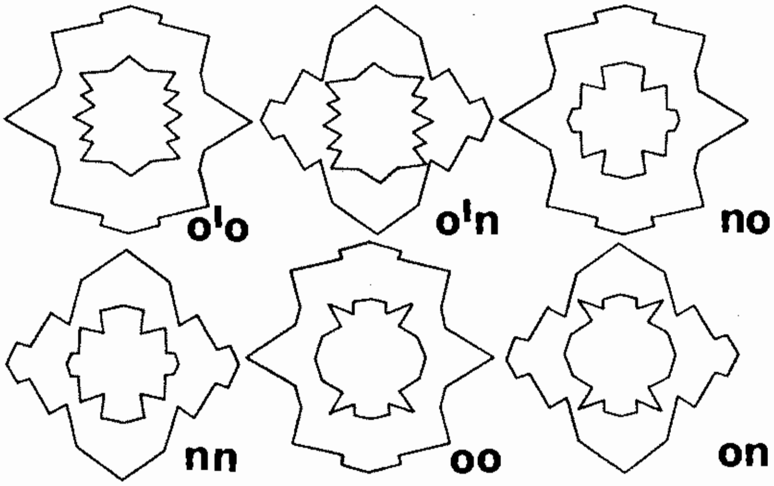
The 24 stimuli and 16 test sets were photographed and reduced to 35 mm negative slides. Three practice stimuli and their two test sets were also drawn and photographed similarly. The practice stimuli were also nested, 24-sided polygons, but care was taken to produce unusual shapes which could easily be seen as "doughnut" shapes. The practice and test stimuli are presented in Appendix 6.

Apparatus

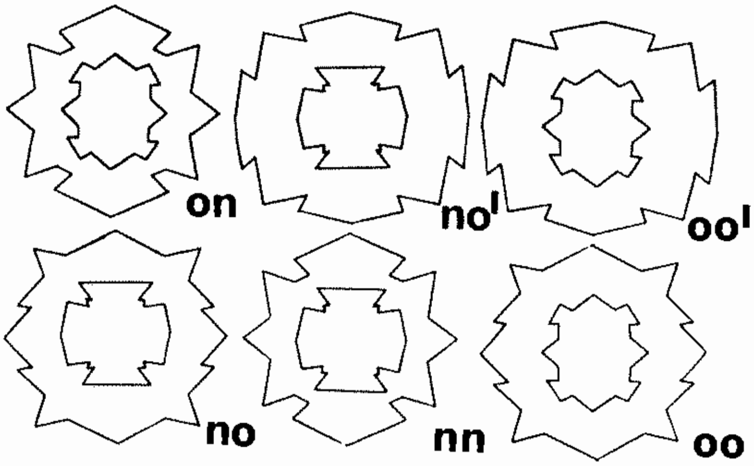
The apparatus and experimental arrangement were those employed in Experiment 7.



Stimulus Set



Test Set 1



Test Set 2

Figure 13. A stimulus set of three shapes, and the two recognition sets which were used to test memory for the first two stimuli, from Experiment 8. The third stimulus (top, right) was not tested, and its parts were used to make the *oo'* and *o'o* distractors. Labels of items in the test sets were not presented during testing.

Procedure

Subjects were seated before the rear projection screen as in the earlier experiments and tested individually, with much the same procedure. One of two types of initial instructions was given to subjects, depending on the group to which they had been assigned.

(1) Figure-Ground Instructions: Subjects in this group were shown a shape similar to those used in testing. This was drawn on a piece of white cardboard. They were told,

> "You can think of (the shape) as a three-dimensional picture, as a shape on top of another shape. The whole thing is sitting on the background."

On another card, the same shape was drawn again, but the inside polygon was darkly shaded and the area between the inside shape the outer polygon was hatched. The figure-ground perception portrayed was discussed with subjects until they said that they could perceive the unshaded picture in the way indicated by the shaded picture. Then subjects were told that they should look at all the pictures to be shown in this way. They were also told that the purpose of the instructions was to determine whether this memory strategy could lead to an improvement in memory for the shapes.

(2) Doughnut Instructions: Subjects in the Doughnut group were told that the special strategy that they were to use was that they should

> "think of the shape as a doughnut. The inside part here is the hole, and the area between the inner and outer boundaries is the doughnut shape sitting on top of the background."

Subjects were given these instructions while shown the same practice shape on cardboard that had been shown to Figure-Ground subjects, plus another version of the shape shaded darkly in the region between the boundaries of the polygons (i.e., with the "doughnut" shaded in). The different figure-ground perception required of subjects in this group was explained carefully to each subject. Some subjects had to be urged not to use their normal figure-ground perceptions.

After the initial instructions, subjects were told that they would be shown three practice shapes and were to attempt to perceive the shapes in the way they had been instructed. The three practice pictures were presented for 0.5 seconds each with 0.5 seconds ISI and approximately five seconds between offset of the last picture and the onset of the first test set. The start of the second test set was paced by the subject, who informed the experimenter when the second set could be shown. Testing of practice shapes and the presentation of the eight lists, each containing three pictures and tested by two recognition sets, was the same as in the studies described in Chapter Seven (including the randomization), except that confidence ratings for inside and outside parts were omitted. Only confidence for whole shapes was rated. This was to discourage Doughnut subjects from perceiving the two parts of each shape as separate configurations. Subjects were asked about their ability to use the perceptual strategy which they had been given after the practice trials and several times during testing, to encourage their consistency in performance.

8.2.2 Results

Means and standard deviations of all responses to items in the test sets, TP scores and CCP scores are presented in Table 32. Contrary to the hypotheses, the Figure-Ground group was somewhat better than the Doughnut group, as evidenced by their higher hit rates, TP and CCP scores. Also, the *nn* rate was much lower for the Figure-Ground group than for the Doughnut group. Figure-Ground subjects were probably guessing on up to 43% of occasions, while Doughnut subjects may have been guessing on 76% of occasions. Thus, Doughnut subjects may have been responding randomly. The most noticeable difference between the two groups was that Figure-Ground subjects chose more *on* than *no* items, while the reverse was true for Doughnut subjects, who appear to have remembered more outsides than insides of shapes.

Table 32

Means and standard deviations of proportions of responses to each item in the test sets, TP scores and CCP scores in each experimental group in Experiment 8.

Score		Instructional Group	
		Figure-Ground	Doughnut
oo	$\bar{X}$	.2548	.1909
	S.D.	.1068	.0943
oo'	$\bar{X}$	.2356	.2466
	S.D.	.1253	.1196
on	$\bar{X}$	.2596	.1690
	S.D.	.1109	.0927
no	$\bar{X}$	.1779	.2672
	S.D.	.0616	.0889
nn	$\bar{X}$	.0721	.1264
	S.D.	.0500	.0883
TP (*oo* + *oo'*)	$\bar{X}$	.4904	.4375
	S.D.	.1345	.1487
CCP $\left(\frac{oo - oo'}{oo + oo'}\right)$	$\bar{X}$	.0545	-0.1756
	S.D.	.4266	0.3917

Hit Rate Analysis

To determine whether the Figure-Ground group was more accurate in recognition than the Doughnut group, mean *oo* scores were compared. No significant difference between these instructional groups was found ($t = 1.5975$, $df = 24$, $p > .05$ for a two-tailed test). Instructions appeared to have no effect on memory for whole stimuli.

The Fragmented Memory Effect

To determine whether subjects in the Figure-Ground group were able to choose correct combinations when they remembered parts of stimuli seen, mean CCP scores for each group were compared in a t-test. There was no significant difference in CCP scores between the two groups ($t = 1.4328$, $df = 24$, $p > .05$ for a two-tailed test). To determine whether subjects in either group were able to discriminate *oo* from *oo'* items, a t-test was carried out on the mean CCP score for each group, to test whether either differed significantly from zero. A t-value of -1.55 was found for the Doughnut group, while a t-value of 0.4427 was found for the Figure-Ground group. Neither of these values is significantly greater than zero in a two-tailed test with 12 degrees of freedom. Subjects could not discriminate between *oo* and *oo'* items in either group.

Memory for One Part of a Shape

Significance tests were carried out on the matched *on* and *no* scores to determine whether subjects in either instructional group remembered one part of the shapes more than another. The Doughnut group had more *no* than *on* responses, but this pattern was reversed for the Figure-Ground group. A t-test carried out on the matched *on* and *no* scores from subjects in each group revealed that there were significantly more *no* than *on* responses in the Doughnut group ($t = 3.0789$, $df = 12$, $p < .01$ for a two-tailed test). The Figure-Ground group made significantly more *on* than *no* responses ($t = 2.2473$, $df = 12$, $p < .05$ for a two-tailed test). It appears that Doughnut subjects were biased towards attending to and remembering outsides of shapes, while Figure-Ground subjects were biased towards attending to and remembering insides of shapes.

At first glance, it appears as though Doughnut subjects were unable to distinguish between *oo*, *oo'* and *no* items, while Figure-Ground subjects could not differentiate between *oo*, *oo'* and *on* items, since proportions of responses to each of these items are similar in the groups. This would suggest that subjects remembered only one part of each shape. However, it must be remembered that *on* and *no* responses occurred twice on half of the test sets, so their chance response rate is 1.5 times that of other response choices. Taking this into account, *oo* and *oo'* responses occur more often than either *on* or *no* responses in each instructional group.

Delay

In a similar group given 0.5 seconds for both presentation time and ISI in Experiment 7, both *oo* and *oo'* scores were affected by the order of presentation of each test set. Replicating the earlier findings, Doughnut subjects had proportionately more *oo* responses on the first test set ($\bar{X} = 0.206$, S.D. = 0.1549) than on the second ($\bar{X} = 0.1758$, S.D. = 0.1327), but the difference was not significant ($t = 0.4966$, $df = 12$, $p > .05$ for a two-tailed test). The mean *oo* scores for the Figure-Ground group were the opposite direction. Subjects made more hits on the second set ($\bar{X} = 0.3077$, S.D. = 0.097) than on the first ($\bar{X} = 0.2019$, S.D. = 0.1806), but again there was no significant difference between these means ($t = 1.9395$, $df = 12$, $p > .05$ for a two-tailed test). Doughnut subjects also chose more *oo'* responses on the first set ($\bar{X} = 0.2692$, S.D. = 0.1828) than on the second ($\bar{X} = 0.2239$, S.D. = 0.1277), but this was not a significant difference ($t = 0.7945$, $df = 12$, $p > .05$ for a two-tailed test). On the other hand, Figure-Ground subjects chose *oo'* items more on the second set tested ($\bar{X} = 0.2404$, S.D. = 0.1077) than the first ($\bar{X} = 0.2308$, S.D. = 0.1755) but the difference was not significant ($t = 0.2336$, $df = 12$, $p > .05$ for a two-tailed test). Thus, order of presentation of the test sets affected neither *oo* nor *oo'* scores, as they had affected scores of a group in Experiment 7 given the same presentation time and ISI as these subjects. The combined effects of delay and interference from the first set did not affect performance on the second recognition set given after each list, for either group.

Confidence Ratings

Confidence ratings are shown in Table 33 for each type of response in both experimental groups. Doughnut subjects gave *nn* items lower ratings than other items, but confidence was equally low for all other response types. These subjects did not think *any* pictures looked very familiar. They gave *oo'* items more positive ratings than any other item type. Confidence was also low in general for the Figure-Ground group, but confidence declined with the increasing number of new parts in a test item, and *oo* items were treated with more confidence than *oo'*

Table 33

Means, frequencies and standard deviations of confidence ratings for test items in each condition in Experiment 8.

Group	Test Item				
	oo	*oo'*	*on*	*no*	*nn*
Figure-Ground					
$\bar{X}$	3.49	3.51	3.59	3.73	3.93
S.D.	1.30	1.14	1.07	1.24	1.03
frequency	53	49	54	37	14
Doughnut					
$\bar{X}$	3.62	3.45	3.51	3.60	4.19
S.D.	1.29	1.46	1.12	1.13	1.20
frequency	39	51	35	55	26

items. These subjects appeared to be more aware of the accuracy of their responses than the Doughnut subjects.

Questionnaires

Post-recognition questionnaires were not very informative in this experiment. Subjects were asked after the experiments whether they were able to use the memory strategy they had been instructed to use and whether they concentrated on any particular section of the shapes. All subjects in the Figure-Ground group reported that they attended primarily to the inside polygons in the shapes. In the Doughnut group, 46% of subjects reported that they attended first, or primarily to the insides, while only 15% reported that they concentrated on the outside parts. An additional 38% of subjects said that they concentrated on either the left or top parts of the entire figure, including portions of both inside and outside parts. Figure-Ground subjects did not report difficulties with their instructions, but Doughnut subjects reported that they found it very difficult to follow instructions about their perceptual strategy. Some subjects reported that they used it only occasionally. Many subjects seemed to be unsure whether they were really perceiving the shapes as "doughnuts".

8.2.3 Discussion

The Instructional Manipulation

The fragmented memory effect, which contributed most of the error observed in the experiments reported in Chapter Seven, was not successfully altered here by instructions which were meant to help subjects to attend to both parts of the stimuli simultaneously. Subjects instructed to perceive two-part shapes as novel figure-ground patterns ("doughnuts") were no more accurate in recognizing correct combinations of parts seen than subjects instructed to perceive the shapes as typical patterns of figure-on-ground. These latter subjects performed the task in a way similar to subjects in Experiments 4 through 7, by attending first to the central figure in the shape and then to the background figure. However, subjects given instructions to perceive the shapes as "doughnuts", to encourage them to attend to both figures in each shape with equal concentration, could not use the instructions to their advantage. Doughnut subjects concentrated instead on outside figures in each shape, and performed no better than the Figure-Ground subjects in the present experiment.

The behaviour of the Doughnut subjects was unusual, since in other circumstances, subjects seem to attend to the figure in a figure-ground shape. For example, subjects viewing a photograph usually make more fixations to the figures than to their surrounds [see examples in Kolers, 1973]. With the two part shapes used in the present experiments, the inner figure seems to have been perceived as the most important part of the shape, judging from subjects' reports. This may be because the inner figure could be seen in its entirety, while the outer figure had the double disadvantage of being partially unseen ("obscured" by the figure resting on top) and acting merely as a background for the central figure. The different attentional pattern implied by Doughnut subjects' performances may have been due to the fact that the important boundary of a "doughnut" is its outer boundary, since this is the contour which carries most information about the concrete object (e.g., this defines its size). If Doughnut subjects altered their usual attentional patterns from attending to the inside part to the outside part, then it is perhaps not surprising that they were no more accurate than Figure-

Ground subjects. Instead of processing global information in the shapes, Doughnut subjects merely substituted a less efficient, focussed attentional strategy for their usual strategy. The concentration on outside parts in the Doughnut group would not be expected to lead to better encoding of the whole shape. There may, in fact, be an advantage in attending to the inside, if only one part can be focussed upon. The advantage of looking at the centre of the screen, in preparation for stimulus presentation, has been mentioned already: the subject does not risk wasting a first eye fixation on a blank part of the screen. A less obvious advantage was pointed out by several subjects in the present experiment. The inner figure is more compact and appears to be a more complex shape because of its closer juxtaposition of lines and angles than the outer figure. The outsides may be perceived as more homogeneous than insides, because their appearance is similar to that of large, distorted circles. Their distinguishing corners are not as discriminable. Of course, a more objective explanation of this may be that the retinal projection of the inner figure is better placed inside the fovea than the outer figure. Both these explanations may be possible.

To reiterate, the Doughnut group's instructions appear to have directed subjects' attention from the inside part of the shapes to the outside, without increasing subjects' tendency to encode information which could link the two parts together. Instead of aiding recognition, the Doughnut strategy impeded accurate encoding by forcing subjects to use an unfamiliar, ineffective strategy. This was evidenced by reports of subjects in the Doughnut group, who said that the task was extremely difficult. The *nn* rate in this group was very high in comparison with that of the Figure-Ground group, and suggests that Doughnut subjects were responding randomly, while Figure-Ground subjects were able to choose correct stimulus parts and their combinations on some occasions. However, the guessing rate in both groups may have been too high to provide an adequate test of memory in this experiment.

Perhaps Doughnut subjects given more practice could use the perceptual strategy to aid their encoding of stimuli. However, there was no evidence that Doughnut subjects improved their recognition accuracy scores towards the end of the testing session. No doubt a longer presentation time and/or ISI would also have aided the performance of Doughnut subjects, but with more time the subjects might have been tempted to use alternative encoding strategies. For this reason, no longer presentation conditions were investigated.

It is surprising that only one subject spontaneously reported noticing that each quadrant of the shapes was the same. Subjects had to remember only one of these quadrants in order to recognize the whole shape. Yet subjects were not able to use this information to increase their recognition accuracy. Perhaps it is not possible to remember disconnected fragments of an entire configuration. This may have been the reason that Doughnut subjects performed so poorly in the present experiment. The shapes seem to have required more than one focussing of attention in order to be remembered, but it may not have been possible for subjects to encode parts of both the inside and outside figures in the same act of attention. Subjects may always attempt to encode only that part of a figure which is complete in some way. This may depend upon Gestalt properties of shapes, such as closure. Thus, there may be principles of structure which direct the subject to organize a picture in a particular way. It may be difficult to design experiments which manipulate the pattern of acts of attention which subjects use to encode a picture. Two studies have offered evidence that some shapes are perceived in a predictable way, with segmentation of the pattern into specific substructures [Reed, 1974; Reed and Johnsen, 1975]. It may be difficult to alter the processing involved in sequential acts of focussed attention in a way that is beneficial to encoding.

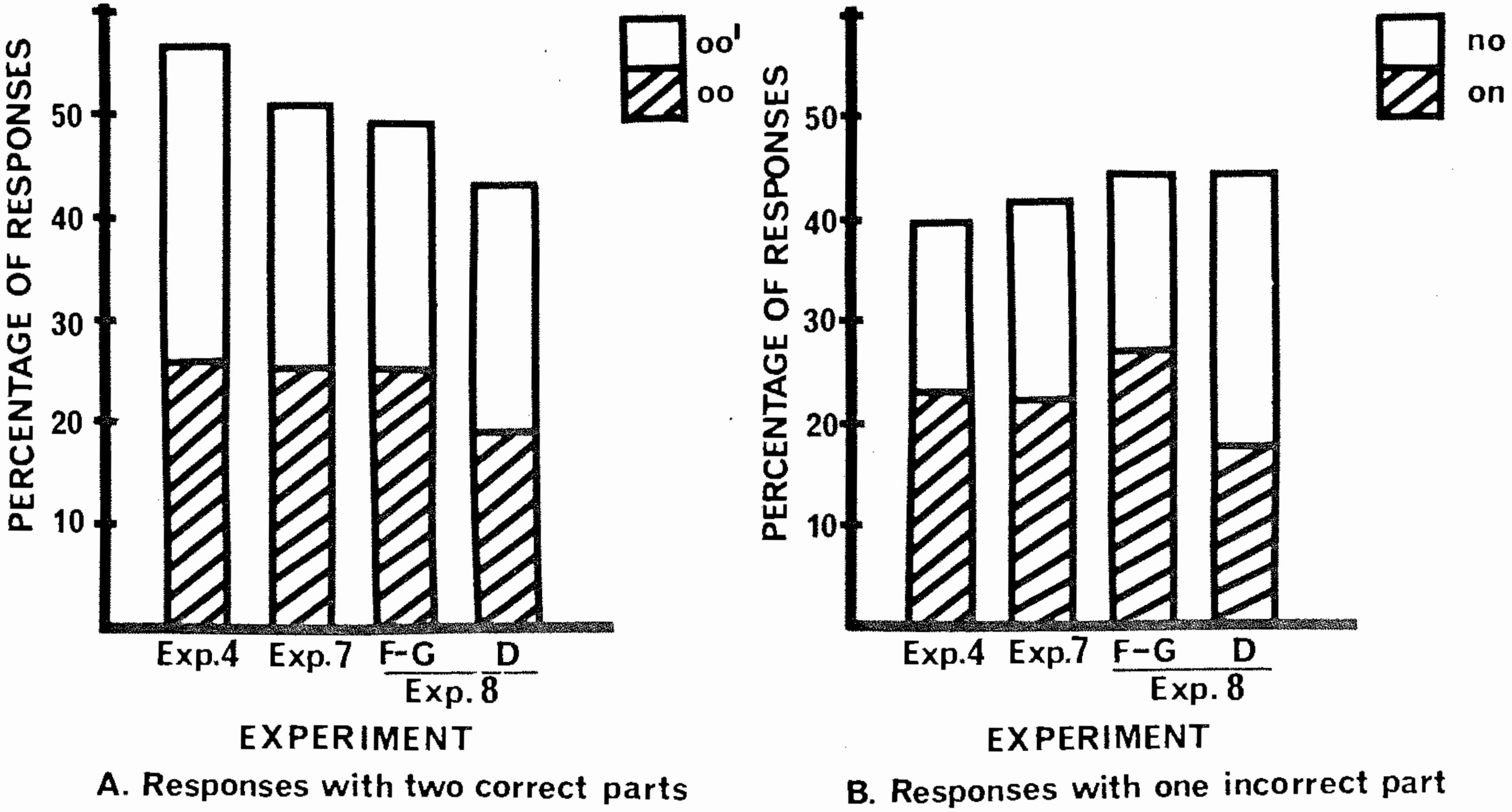
The unpredictable effects of manipulations of strategies in the present experiment and in Experiments 2 and 3 using the exclusion set paradigm may have been due to difficulties in altering attentional patterns. From the studies reported here, it appears that the manipulation of the instructions to alter attention to a stimulus is a procedure of doubtful reliability.

Experiment 8 as a Control for Experiments in Chapter Seven

A comparison of the results of the two groups in the present experiment with results from the 0.5/0.5 groups in Experiments 4 and 7

(Chapter Seven) provides further evidence concerning the processing involved at fast presentation rates. Distributions of responses for each recognition test item in the three groups are shown in Figure 14. The performance of Figure-Ground subjects in the present experiment is very similar to that of the group in Experiment 7 in its proportions of *oo* and *oo'* responses. The groups differ, however, in their patterns of *on* and *no* responses. Though subjects in both groups reported a bias towards attending to insides rather than outsides of their respective stimuli, only the Figure-Ground group in the present experiment was biased in memory scores so that more insides than outsides of shapes were remembered. This group provides a control for the experimental groups in Chapter Seven. Subjects' biases in attending to central figures in these experiments could have been due either to differences in attentional patterns or to the use of different types of configuration. The results of the Figure-Ground group suggest that the bias towards central attending is a normal encoding tendency, and is not due to uncontrolled factors which might differentiate between inside and outside patterns (except size). However, there was also evidence from subjects given Doughnut instructions that this tendency could be altered, from an attention to insides to an attention to outsides.

It is interesting that the performance of the Figure-Ground group in the present experiment is more similar to the performance of the 0.5/0.5 group in Experiment 4 than to the equivalent group in Experiment 7. The subjects in Experiment 4, like Figure-Ground subjects here, remembered more *on* than *no* items. In Experiment 4, this was presumed to be due to inadequate time to encode more than one part of the stimulus, but clearly this is not the case. On many occasions, subjects in the 0.5/0.5 group in Experiment 7 were able to remember both parts of the shapes, and their choices of test items containing only one correct part occurred no more than would have been expected by chance. Perhaps the normal bias in subjects' attentional strategies only affects memory when the recognition task is very difficult. The recognition task in Experiment 7 may have been easier for subjects, allowing them to demonstrate their accurate memories for both insides and outsides of shapes. Of course, it is difficult from these results to determine whether subjects given a difficult task actually remember only one part



A. Responses with two correct parts

B. Responses with one incorrect part

Figure 14. Percentages of: (A) both *oo* and *oo'* (including *o'o*) responses, and (B) both *on* (including *o'n*) and *no* (including *no'*) responses, for both experimental groups given 0.5 seconds of presentation time (PT) and interstimulus interval (ISI) in Experiments 4 and 7, and for the Figure-Ground (F-G) and Doughnut (D) groups in Experiment 8.

of each shape or merely find the task of discriminating between old and new parts more difficult for the outsides (or insides). Data from the Figure-Ground group in the present experiment shows that subjects remember only one part of a shape on many occasions when the difficulty of the recognition choice is the same for both inside and outside parts, i.e., when both parts are the same time of configuration. Thus, there seems adequate evidence to suggest that, when subjects find a recognition task very difficult, they resort to an attentional strategy which ensures that they remember at least one part of each stimulus. Under normal viewing conditions, both parts are encoded, though the part of a picture usually given most emphasis in attention appears to be the figure in a picture with figure-ground contrast.

Conclusions about the Fragmented Memory Effect

None of the manipulations of temporal variables or perceptual strategy were able to produce changes in the measure of fragmented memory (CCP scores). However, since the scores were based on the relationship between the hit rate (*OO* responses) and the rate of choosing distractors made of recombined parts of different stimuli (*OO'* responses), the significant changes in the hit rate due to experimental conditions should have led to corresponding changes in CCP scores. A common problem in significance testing based on probability estimates is that results of various comparisons may be inconsistent with each other. In new fields of research, it is not always possible to make firm predictions about the outcomes of experiments in order to make orthogonal comparisons. Thus, though in the present studies there was a trend towards higher CCP scores when subjects were given more time to process each stimulus, this cannot be used as evidence supporting the view that temporal variables affect fragmented memory.

However, these studies do provide strong evidence that there is a fragmented memory effect, since *OO'* distractors were chosen much more often than any other distractors and almost as often as the *OO* items. Even in groups in which subjects were given ample time to encode each stimulus, subjects were often unable to remember how parts were combined in the pictures, though they had very accurate memories for each individual part. This suggests that the encoding process may be disrupted so that events are not adequately encoded into memory. The study of this phenomenon may lead to a better understanding of integrative processes which occur in remembering. It may be that the fragmented memory effect can be manipulated by using different experimental conditions. This deserves further investigation.

CHAPTER NINE

SUMMARY AND CONCLUDING REMARKS

9.1 SUMMARY OF FINDINGS

Most conceptualizations of memory are based on empirical evidence from studies employing only verbal stimuli. If any general theory of memory is to be developed, there must be broader experimental investigations using different types of stimuli and methodologies. The general aim of this research was to examine the characteristics of the remembering process when pictorial stimuli are presented. The pictorial stimuli used were both realistic, representational drawings and abstract shapes.

The review of past findings in Chapter Two showed that no available method for testing memory for pictures overcomes the problem that experimental subjects cannot communicate their memories for visual, nonverbal material adequately. Recognition methods seem to be preferable to recall, since the former are not influenced either by subjects' production difficulties or their verbal fluency, and they yield unambiguous quantitative scores. However, traditional recognition paradigms also have drawbacks. Subjects usually have only to choose an item from a set of test items, and they may choose the correct stimulus by chance. Also, since pictures in the test sets may differ from each other in many ways, subjects may be able to reject all the incorrect items on the basis of only partial remembered information about the stimulus. The recognition response only indicates whether subjects have remembered something about each picture. Thus, the traditional recognition task fails to allow subjects to be given discriminative accuracy scores for each stimulus. Because of this, the high estimates of recognition accuracy in past studies may have been spuriously inflated, leading to the view that memory for pictures has phenomenally high capacity, and that stimuli are registered automatically in memory as quasi-photographic traces.

The experiments described were mainly directed towards examining this view of the pictorial memory process. Evidence reported suggests that memory for pictures is in fact limited in capacity and that the encoding of pictures is a complex activity under the subject's control.

9.1.1 Experiment 1: The Capacity of Memory for Pictures

A new recognition paradigm was developed for testing memory for pictures, the exclusion set method (see section 2.5.2), in order to overcome some of the disadvantages of traditional methods. This method yielded an accuracy score which quantified each subject's memory for detail in a single picture. The method was successfully employed to measure the decline in recognition accuracy of four complex, representational drawings over a two-month period in Experiment 1. There was no decay in memory during the first two days after the pictures were presented, but recognition accuracy fell significantly after one week and again after two months. This suggests that memory for pictures is not quasi-photographic, but is limited in capacity. The gradual increase in forgetting over time also suggests that there is not the obvious dichotomy between recognition accuracy in the immediate testing situation and the delayed-testing situation in memory for pictures which is found in memory for words. Perhaps this means that memory for pictures does have unusually stable memory traces. However, differences in the methodologies of experiments testing memory for pictures and words make it impossible to compare their results. The findings do suggest that there may not be the limited-capacity short-term memory for pictures which has often been postulated for words. Instead, memory capacity for detailed information seems to be fairly accurate when tested immediately and declines only slowly over time.

9.1.2 Experiments 2 and 3: Effects of Intentional and Incidental Instructions on Memory for Pictures

The exclusion set recognition method was also employed in Experiments 2 and 3 to test the hypothesis that pictures are encoded into memory automatically, regardless of the subject's activities during encoding. This hypothesis predicts that subjects motivated to notice small details in each picture in order to remember them will not be any more accurate in recognition than subjects who have only glanced at the pictures without observing details carefully.

Studies of recognition of words have traditionally found that subjects do not have to be motivated to remember words for the recognition task, as the same amount of information is remembered when the subject is prepared for the recognition task as when the subject attends to the words but is unaware that he must try to remember them. This finding is consistent with the automatic encoding hypothesis. To test the generality of the findings from verbal studies, there was a traditional manipulation of intentional and incidental strategies in Experiment 2. One group of subjects (the Incidental group) was given no instructions about the exclusion set task, and instead was given an orienting task of classifying items in the pictures as a "blind" to encourage the subjects to attend to each picture. These subjects recognized pictures as accurately as subjects given the orienting task who were told that the stimuli would be tested for recognition (the Control group). The orienting task itself apparently had no effect on subjects' encoding of pictures, since subjects in a standard Intentional group prepared for the recognition test and allowed to view the pictures freely without an orienting task performed no better than Control subjects. The results of this experiment seemed to support the automatic encoding hypothesis.

However, there was a modified replication of this study in Experiment 3. The Intentional group in this experiment was given a practice trial in place of unrestricted viewing of the pictures. These subjects *did* remember more about the stimuli than subjects given an orienting task of counting circles in the pictures. However, there was again no difference found between Incidental and Control groups given the orienting task and differing only in whether they were told about the memory task. Thus, subjects given practice with the recognition task could remember more about the stimuli than subjects merely told about the recognition task and asked to carry out an orienting task activity while viewing the pictures.

It was found that the superiority of the Intentional subjects' performance in Experiment 3 was not due simply to the fact that they were given an extra trial on which to learn strategies for carrying out the recognition task. The results suggest that certain activities carried out by the subject during encoding may lead to more accurate memory than other activities. Intentional subjects appear to have noticed more details in each stimulus, because they had been given a difficult discrimination task (the practice trial) before list presentation. They may have been cued to encode as much detail as possible from the pictures. Apparently, subjects given the orienting task were not able to carry out encoding processes with this degree of elaboration during stimulus presentation. This suggests that encoding of pictures is an active, variable process under the subject's direct control. Since subjects can be encouraged to encode more detail from each picture when they are given appropriate instructions, encoding cannot be an automatic, passive process.

Experiments 2 and 3 also offered evidence refuting the hypothesis that motivation alone is important in recognition of pictures when a difficult task is used to test recognition. Subjects given orienting instructions did not encode any more information from pictures when they were aware that their memories were to be tested (and presumably when they were motivated to try to commit the stimuli to memory) than when they were unprepared for the memory task. These subjects may have expended the same effort in attending to stimuli regardless of their motivation. The most important variable leading to memory accuracy appears to be the subject's level of attention to details in the pictures. Stimuli given little attention will not be encoded as well as pictures given more attention to detail during encoding.

9.1.3 Experiments 4 to 7: Effects of Temporal Parameters on Memory for Shapes

Since memory for pictures does not appear to be automatic, an obvious strategy for examining the encoding process itself would be to examine its temporal stages. This would determine the length of time necessary for adequate encoding. The studies reported in Chapter Seven were aimed at discovering whether the integration of parts attended to separately during encoding is automatic, or whether subjects require more time to be able to link parts seen together in memory. One interesting problem not given much consideration in previous studies is the existence of a fragmented memory effect. There is evidence that subjects can make confusions between pictures when parts are remembered from a scene, suggesting that adequate encoding of parts does not ensure that the whole has been registered in memory. It was thought that, if the act of registering "whole events" in memory is not automatic, this effect might be more likely to occur when subjects are not given adequate time for encoding.

For the experiments reported in Chapter Seven, a recognition task was devised to test memory for parts of shapes as well as their combinations. This was used to estimate the extent of fragmented memory. In Experiment 4, the method was used to test recognition accuracy when both presentation time and ISI were varied between 0.5 and two seconds. A large fragmented memory effect was found, since subjects could remember parts of the shapes seen but could not distinguish between correct stimuli and distractors made by recombining parts of different stimuli. However, there was no effect of either presentation time or ISI on recognition accuracy. This failed to replicate findings of previous investigators, and appeared to be due to the difficulty of the recognition task. In Experiment 5, the methodology was altered so that stimuli and distractors were less confusable. Longer exposures were given for each stimulus to attempt to increase recognition accuracy. The experiment found that there was an increase in the hit rate as presentation time increased from two to five seconds, with ISI held constant at 1.5 seconds. There was still evidence of a large fragmented memory effect, but subjects in this experiment were able to choose test items on most occasions when both parts of a shape were remembered, regardless of their presentation time condition.

In Experiment 6, only ISI was varied. No increase in either recognition accuracy or the fragmented memory effect was found when ISI was increased from 1.5 to seven seconds, with presentation time held constant at two seconds. However, a partial replication of Experiment 4 using the new test sets did find an effect of increasing ISI on recognition accuracy. Experiment 7 showed that recognition accuracy increased as both presentation time and ISI were increased from 0.5 to two seconds. The most important temporal factor was found to be presentation time, since, given the same total time between onset of one picture and onset of the next in the list, a longer presentation time was more beneficial than a longer ISI. This refuted the total time hypothesis. Subjects in this experiment were able to remember both parts of the shapes equally well, but they exhibited a large fragmented memory effect. The effect was larger in this experiment than in Experiment 6, presumably because when subjects are given less total time to view each picture, they are not as able to remember how the parts were combined. The results of this study failed to replicate those of Experiment 4, suggesting that methodological problems in the latter had made the task too difficult for subjects.

The results of Experiments 5, 6 and 7, taken together, suggest that more information can be encoded from a picture the longer it is available for viewing. This provides further evidence that encoding of pictures is influenced by the elaboration of the subject's activities. The results also suggested that the blank time between stimuli can be used to benefit memory processing. Subjects may still be able to extract information from the picture (e.g., from an icon) once it has been removed from view. It seems possible that subjects can use the ISI to organize information previously extracted from the picture or to rehearse what was encoded. On the other hand, the longer ISI may allow more encoded information to be consolidated into stable memory traces before there is interference from the next stimulus, or it may simply give each event a more unique temporal context. The extent of the subject's control over activities carried out during the ISI needs further investigation.

The results of these experiments are consistent with the view that attention proceeds by a series of operations of focussed concentration upon subareas of the whole picture, followed by some integration of these memories for separate parts into a memory for the whole configuration. These focussed attentions may be made after some global information (a "general impression") has been extracted from the picture. However, the presence of the fragmented memory effect suggests that any initial global information extracted from the pictures is often inadequate for tying all the details together in memory. It was suggested in Chapter Eight that a final, global analysis of the picture may be necessary for complete encoding of a picture as an event. This could take the form of an act of attention to the picture after all the important details have been attended to individually. In this final stage of attention, there could be constructive encoding of relationships between separate features of the picture. Perhaps there is some abbreviated cue stored at this stage, which can be used later to retrieve the entire structural description of the picture [cf., Sutherland, 1968] from memory.

9.1.4 Experiment 8: Effects of Perceptual Strategy on Recognition Memory

The stimuli used in Experiments 4 to 7 were simple figures usually perceived as figure-ground patterns. There was evidence from confidence ratings and subjects' reports that the central figure in each shape was easier to remember than its surrounding background shape. In Experiment 8, the fast (0.5 second) presentation conditions of Experiments 4 and 7 were replicated with new stimuli, which were controlled so that the central figures and outsides were equal in difficulty. It was hypothesized that subjects attending to the shapes as figure-ground patterns (Figure-Ground subjects) would perform similarly to subjects in comparable groups in Experiments 4 and 7, who reported that this was their usual perceptual strategy. Another group of subjects was included, which was instructed to perceive a different figure-ground organization of the shapes (Doughnut subjects), the contours of both the inside and outside figures being perceived as the boundaries of a single surface. It was predicted that these subjects would not attend first to the inside figure and then to the outside, like subjects in the other groups, but would attend to both parts simultaneously. If this were true, it would be expected that the fragmented memory effect would not be as evident in their performance, since they would not have encoded each part separately in memory.

Neither hypothesis was given full support in this experiment. Subjects in the Doughnut group appeared to find the task very difficult. The groups did not differ in recognition accuracy or in frequencies of recognition responses affected by fragmented memory. However, there was evidence that subjects in the two groups used different encoding activities, since Figure-Ground subjects remembered more insides than outsides of shapes, while Doughnut subjects remembered more outsides than insides. This suggested that subjects normally attend first to the inside of a figure-ground shape, but that this tendency can be altered with appropriate instructions. This provides further evidence that encoding activities are under the subject's control. Since subjects in Experiment 7 could remember both parts of stimuli shown on most occasions, even in the very fast presentation conditions, it seems that 0.5 seconds is enough time to encode both parts of the shapes. However, subjects may *choose* to encode only one part if the recognition task is very difficult. This suggests that subjects' acts of attention are not of fixed duration and information-handling capacity, and argues against the eye-fixation analogy of attentional scanning. Encoding is a variable process which the subject can manipulate to his advantage. This is consistent with depth-of-processing views of memory [cf., Craik and Lockhart, 1972].

These experiments offered no evidence that only subareas of pictures are encoded in each focussing of attention. To demonstrate this, faster presentation times would have to be used to investigate whether, at some brief presentation time, subjects can encode a small amount of detail about a picture but not information about all the parts. This was not possible given the recognition testing method used in the present experiment. A more appropriate method for testing this notion might be a recognition task which tests memory for a single item at a time, in order to minimize the delay before testing which affects memory for briefly presented visual delays.

The experiments reported in Chapters Seven and Eight demonstrated the existence of a fragmented memory effect. It appears as though, in a difficult recognition task like the present one, subjects may often remember information about the parts of stimuli seen and yet fail to encode any information about their combinations. There was evidence that subjects given only 0.5 seconds presentation time to view each picture had no information about combinations, though they were quite accurate in remembering both parts of pictures seen. Subjects given more viewing time (two or five seconds) were more likely to be able to choose correct combinations, given that they could remember parts seen. On the other hand, there was no clear evidence that lengthening ISI aided recognition and led to less effect of fragmented memory. These tentative results suggest that, in order to encode the parts remembered as a whole, subjects must have time to view the stimulus after all the details from the subareas of the picture have been encoded.

9.2 CONCLUDING REMARKS

9.2.1 Implications for a Theory of Memory

The research reported here may be seen as broadening the approach to the study of memory. Any conceptualization of a general memory process must take into account findings of experiments using diverse methodologies in order to separate general characteristics of the remembering process from those specific to a narrow range of stimuli or experimental situations.

There was evidence in Experiment 1 that memory for representational pictures, like memory for words, decays over time. However, there may be a more gradual decline for pictorial than verbal stimuli. Memory for pictures appears to be as accurate when tested immediately as when tested two days later. Since memory for words, when testing is carried out immediately, appears to be quite different from memory when testing is delayed, many models of memory have postulated that there is a limited-capacity short-term memory store which is separate from a long-term store. Evidence from studies measuring memory for pictures does not support the notion that this dichotomy is a general property of memory. Short delays affect neither memory for representational drawings nor memory for shapes, as long as subjects have adequate time to encode the pictures during their presentation. (This does not include the very short delays studied by Posner [1969] in a different type of recognition methodology.)

The hypothesis that encoding of pictures is an automatic process which leads to perfect retrieval of the information in a recognition task was not given support in this study. Memory for pictures can be very accurate, even over long periods of time, but this is a function of the elaboration of the subject's initial encoding processes while viewing the stimulus. Though subjects' motivation to remember pictures by itself does not appear to lead to more efficient encoding, subjects do encode more information from pictures and recognize them more accurately when they are cued to attend to minute details in each picture during its presentation.

The results of this research are consistent with the view that attentional scanning is a sequential process. It can be supposed that the subject's initial act of attending to a picture yields a "general impression" which guides the subject's organization of the pattern into substructures. Then, a sequence of individual acts of focussed attention to the substructures may be carried out to encode a structural description of the pattern. These "acts" are probably not of finite duration and information-handling capacity, since subjects can be instructed to vary their patterns of attention to more or less information in each visual array. Subjects seem to be able to vary the order in which parts of the picture are attended to, in order to maximize the amount of information extracted, given the demands of the situation. This suggests that encoding is a constructive process guided actively by the subject.

The ubiquitous presence of the fragmented memory effect suggests that often subjects' encoding of events may be inadequate. Even when subjects have ample time to encode all the detail in a picture in a series of acts of focussed attention, parts may still be disconnected in memory. This problem deserves further investigation. Perhaps, with more time, subjects are more able to notice relationships between substructures of the picture in order to link them together in memory. This may require further acts of attention *after* each substructure has been encoded in memory. For example, subjects might use further acts of attention to notice how the parts are related to each other. On the other hand, they may recode the pattern into different substructures so that there is more cross-classification of minute details in various acts of attention. Some pictures, like the two-part shapes used here, may be perceived with a stereotypical perceptual organization so that parts are encoded separately in memory with little information about the context in which they were perceived. Other stimuli may allow subjects to attend to a single area of the picture which contains information overlapping with information attended to in another act of attention. The fragmented memory effect may be a function of the amount of overlap in information about details in the picture encoded in each attentional act. This overlap may be maximized in representational drawings and minimized in abstract shapes.

The fragmented memory effect may also be just a passive outcome of overloading of the encoding processes' organizational capacity. Perhaps under most conditions, events are separated in time or there are many distinctive cues which enable any item of information remembered to be put into a context. The experimental situation devised to demonstrate the fragmented memory effect may have disrupted a fairly automatic process of integration of parts into a whole which occurs in normal encoding. It would be interesting to investigate whether subjects given a list of pictures in one experimental setting and given another list in a different environment (e.g., another room) would make as many inter-list confusions as intra-list confusions. The role of the total experimental context in aiding subjects' encoding and retrieval of information should be investigated, since this must be of ultimate importance in guiding the subjects' responses during the memory task.

9.2.2 Recognition Methodology: Future Directions

Recall methods of testing memory for pictures were found to be unsatisfactory, because it is difficult to determine the subject's accuracy of memory on the basis of his drawings or verbal descriptions. However, studies using the reproduction method for testing memory have yielded much richer descriptions of errors which occur in forgetting than have studies employing recognition paradigms. This suggests that recognition can only measure very specialized types of memory loss. Because this usually makes the recognition task easier than a recall task, it is often assumed that the process of recognizing a picture is a simple, one-step retrieval operation, while recall is a more elaborated search of memory and retrieval. However, until more is known about the process of recognition itself, the method cannot be discarded as a measure of only the most simple of memory processes.

Traditional recognition tasks have, in general, been easy tasks for subjects, in which they have been able to rely on partial information about the stimuli to achieve accurate scores. However, this may not be a general property of all recognition situations. The results of the experiments reported here using the exclusion set method show that recognition may be a very difficult task, in which only the most accurate subjects can use information about a single part of a picture in order to achieve a high recognition score. This suggests that subjects must use appropriate encoding strategies in order to remember the pictures and retrieve fairly detailed information from memory about the stimuli. Subjects did not seem to be able to accept or reject a picture immediately on the basis of its familiarity; the decision probably required a search of stored information.

Though the exclusion set recognition task was difficult, it was used only to measure the amount of detail which subjects had lost from memory. In Experiments 1 and 2, loss of various kinds of information was sampled in the distractor sets, but in Experiment 3, only loss of positional information about items in the array was tested. This was done in order to have greater control over the similarity of items in the test sets and consequently of the type of information lost from memory. It might be a fruitful approach to investigate losses of different kinds of information separately in future studies, before there is any attempt to examine the effects of their interactions. For example, the experimental conditions which lead to the subject's loss of positional information may be different to those which lead to forgetting of the structural information in a geometric form.

The exclusion set method was not used to test any errors in memory except loss of information. The studies of subjects' reproductions of pictures from memory (see section 2.1) suggested that there are many other kinds of distortion and loss of memory which deserve more thorough study. One of these, interpolation or confusion between pictures, was successfully demonstrated in the recognition task employed in Experiments 4 to 8. It may be possible to test other kinds of forgetting with new recognition paradigms. For example, one of the most interesting errors was subjects' confabulation of remembered detail to produce a drawing, when they seemed to be able to remember only disconnected fragments of visual detail without remembering much semantic information about items in the pictures or global aspects of scenes portrayed. This error would be extremely difficult to test in a recognition task, because it is difficult to create distractors which would foil subjects with idiosyncratic memories. Perhaps with very simple representational pictures the possibilities of subjects' confabulations could be limited in order to produce distractors that subjects with confabulated memories might falsely recognize. This might be a fruitful direction for future research in recognition.

The exclusion set method also has untapped potential for measuring recognition of other types of material, because the principles with which it was developed can be extended to any situation in which the subject is assumed to be able to exclude stimuli as unfamiliar when they contain some feature which is consistent with the subject's description of the stimulus in memory. All that is required is that recognition distractors be created by varying the stimulus item on some dimension. For example, arrays of moving objects on film could be tested in a recognition task with distractors made from similar films in which displacement of objects in space or the velocity of their movement could be varied. The method might also open the way to controlled investigations of memory for three-dimensional objects, by varying the spatial dimensions of an object to create its distractors in a recognition test. This could broaden the scope of experimental investigations of memory processes, which might eventually lead to the construction of a general theory of memory.

REFERENCES

Allport, G. Change and decay in the visual memory image. *British Journal of Psychology*, 1930, 21, 133-148.

Arnoult, M.D. Familiarity and recognition of nonsense shapes. *Journal of Experimental Psychology*, 1956, 51, 269-276.

Atkinson, R.C. and Shiffrin, R.M. Human memory: a proposed system and its control processes. In K.W. Spence and J.T. Spence (Eds.), *Advances in the psychology of learning and motivation*, Vol. 2, New York: Academic Press, 1968.

Attneave, F. Symmetry, information and memory for patterns. *American Journal of Psychology*, 1955, 68, 209-222.

Attneave, F. *Applications of information theory to psychology.* New York: Holt, Rinehart and Winston, 1959.

Bahrick, H.P. Retention curves: facts or artifacts? *Psychological Bulletin*, 1964, 61, 188-194.

Bahrick, H.P. and Bahrick, P. Independency of verbal and visual codes of the same stimuli. *Journal of Experimental Psychology*, 1971, 91, 344-346.

Bahrick, H.P. and Boucher, B. Retention of visual and verbal codes of the same stimuli. *Journal of Experimental Psychology*, 1968, 78, 417-422.

Bahrick, H.P., Clark, S. and Bahrick, P. Generalization gradients as indicants of learning and retention of a recognition task. *Journal of Experimental Psychology*, 1967, 75, 464-471.

Bartlett, F.C. *Remembering.* Cambridge: Cambridge University Press, 1932.

Benton, A.L. and Van Allen, M.W. Impairment of facial recognition in patients with a cerebral disease. *Cortex*, 1970, 4, 344-358.

Bousfield, W.A., Berkowitz, H. and Whitmarsh, G. Association clustering in the recall of minimally meaningful geometric designs. *Canadian Journal of Psychology*, 1959, 13, 281-287.

Bower, G.H. and Karlin, M.B. Depth of processing pictures of faces and recognition memory. *Journal of Experimental Psychology*, 1974, 103, 751-757.

Box, G.E.P. Non-normality and tests on variance. *Biometrika*, 1953, 40, 318-335.

Broadbent, D.E. *Perception and communication.* New York: Pergamon Press, 1958.

Brown, A.L. and Campione, J.C. Recognition memory for perceptually similar pictures in preschool children. *Journal of Experimental Psychology*, 1972, 95, 55-62.

Brown, A.L. and Scott, M.S. Recognition memory for pictures in preschool children. *Journal of Experimental Child Psychology*, 1971, 11, 401-412.

Brown, R.W. and Lenneberg, E.H. A study in language and cognition. *Journal of Abnormal and Social Psychology*, 1954, 49, 454-462.

Bruner, J.S., Busiek, R. and Minturn, A. Assimilation in the immediate reproduction of visually perceived figures. *Journal of Experimental Psychology*, 1952, 44, 151-155.

Carmichael, L., Hogan, H.P. and Walter, A.A. An experimental study of the effect of language on the reproduction of visually perceived form. *Journal of Experimental Psychology*, 1932, 15, 73-86.

Cermak, G.W. Short-term recognition memory for complex free-form figures. *Psychonomic Science*, 1971, 25, 209-211.

Clark, H.J. Recognition memory for random shapes as a function of complexity, association value, and delay. *Journal of Experimental Psychology*, 1965, 69, 590-595.

Clark, H.J. Random shape recognition at brief exposure durations. *Psychonomic Science*, 1968, 12, 245-246.

Cohen, G. Serial position effects in the recall of picture sequences. *Quarterly Journal of Experimental Psychology*, 1972, 24, 41-47.

Cohen, R.L. Effect of verbal labels on recall of a visually perceived simple figure: recognition versus reproduction. *Perceptual and Motor Skills*, 1966, 23, 859-862.

Cohen, R.L. and Granstrom, K. Interpolated task and mode of recall as variables in short-term memory for visual figures. *Journal of Verbal Learning and Verbal Behavior*, 1968, 7, 653-658.

Coltheart, M. Visual information-processing. In E. Dodwell (Ed.), *New horizons in psychology*, Vol. 2. Middlesex: Penguin, 1972.

Craik, F.I.M. A "levels of analysis" view of memory. In P. Pliner, L. Krames and T.M. Alloway (Eds.), *Communication and affect: language and thought.* New York: Academic Press, 1973.

Craik, F.I.M. and Lockhart, R.S. Levels of processing: a framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 1972, 11, 671-684.

Craik, F.I.M. and Tulving, E. Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, 1975, 104, 268-294.

Cross, J.F., Cross, J. and Daly, J. Sex, race, age, and beauty as factors in recognition of faces. *Perception and Psychophysics*, 1971, 10, 393-396.

Dallett, K. and Wilcox, S.G. Remembering pictures versus remembering descriptions. *Psychonomic Science*, 1968, 11, 139-140.

Daniel, T.C. Nature of the effect of verbal labels on recognition memory for forms. *Journal of Experimental Psychology*, 1972, 96, 152-157.

Daniel, T.C. and Ellis, H.C. Stimulus codability and long-term recognition memory for visual form. *Journal of Experimental Psychology*, 1972, 93, 83-89.

Deese, J. Influence of inter-item associative strength upon immediate free recall. *Psychological Reports*, 1959, 5, 305-312.

Dornbush, R.L. and Winnick, W.A. Short-term intentional and incidental learning. *Journal of Experimental Psychology*, 1967, 73, 608-611.

Eagle, M. and Leiter, E. Recall and recognition in intentional and incidental learning. *Journal of Experimental Psychology*, 1964, 68, 58-63.

Ebbinghaus, H. *Memory: A contribution to experimental psychology*. New York: Dover, 1964. (Originally published: Leipzig: Duncker and Humblot, 1885.)

Ellis, H.C. Transfer of stimulus predifferentiation to shape recognition and identification learning: Role of properties of verbal labels. *Journal of Experimental Psychology*, 1968, 78, 401-409.

Ellis, H.C. and Daniel, T.C. Verbal processes in long-term stimulus recognition memory. *Journal of Experimental Psychology*, 1971, 90, 18-26.

Ellis, H.C. and Homan, L.E. Implicit verbal responses and the transfer of stimulus predifferentiation. *Journal of Experimental Psychology*, 1968, 76, 486-489.

Epstein, M.L., Phillips, W.D. and Johnson, S.J. Recall of related and unrelated word pairs as a function of processing level. *Journal of Experimental Psychology*, 1975, 104, 149-152.

Fischler, I. and Puff, C.R. Organization in free recall with verbal and pictorial modes of input and output. *Psychonomic Science*, 1971, 22, 85-86.

Franken, R.E. and Davis, J. Predicting memory for pictures from rankings of interestingness, pleasingness, complexity, figure-ground and clarity. *Perceptual and Motor Skills*, 1975, 41, 243-247.

Friedman, M.P., Reed, S.K. and Carterette, E.C. Feature saliency and recognition memory for schematic faces. *Perception and Psychophysics*, 1971, 10, 47-50.

Frijda, N. and Van de Geer, J.P. Codability and recognition. *Acta Psychologica*, 1961, 18, 360-367.

Frost, N. Clustering by visual shape in the free recall of pictorial stimuli. *Journal of Experimental Psychology*, 1971, 88, 409-413.

Frost, N. Encoding and retrieval in visual memory tasks. *Journal of Experimental Psychology*, 1972, 95, 317-326.

Galper, R.E. Recognition of faces in the photographic negative. *Psychonomic Science*, 1970, 19, 207-208.

Galper, R.E. and Hochberg, J. Recognition memory for photographs of faces. *American Journal of Psychology*, 1971, 84, 351-354.

Gibson, J.J. The reproduction of visually perceived forms. *Journal of Experimental Psychology*, 1929, 13, 1-39.

Gibson, J.J. Pictures, perspective and perception. *Daedalus*, 1960, 89, 216-227.

Glanzer, M. and Clark, W.H. Accuracy of perceptual recall: an analysis of organization. *Journal of Verbal Learning and Verbal Behavior*, 1963, 1, 289-299.

Glanzer, M. and Clark, W.H. The verbal loop hypothesis: conventional figures. *American Journal of Psychology*, 1964, 77, 621-626.

Glanzer, M. and Cunitz, A.R. Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior*, 1966, 5, 351-360.

Goldstein, A.G. Learning of inverted and normally oriented faces in children and adults. *Phsychonomic Science*, 1965, 3, 447-448.

Goldstein, A.G. and Chance, J.E. Visual recognition memory for complex configurations. *Perception and Psychophysics*, 1970, 9, 237-241.

Griffith, D. Comparison of control processes for recognition and recall. *Journal of Experimental Psychology: Human Learning and Memory*, 1975, 104, 223-228.

Haith, M.M. Developmental changes in visual information processing and short-term visual memory. *Human Development*, 1971, 14, 249-261.

Hanawalt, N.G. and Demarest, I.H. The effect of verbal suggestion in the recall period upon the reproduction of visually perceived forms. *Journal of Experimental Psychology*, 1939, 25, 159-174.

Hebb, D.O. The American Revolution. Address of the President to the 68th Annual Convention of the American Psychological Association. Chicago, September 4, 1960.

Herman, D.T., Lawless, R.H. and Marshall, R.W. Variables in the effect of language on the reproduction of visually perceived forms. *Perceptual and Motor Skills*, 1957, 7, 171-186. Monograph Supplement #2.

Hines, D. Immediate and delayed recognition of sequentially presented random shapes. *Journal of Experimental Psychology: Human Learning and Memory*, 1975, 1, 634-639.

Hintzman, D.L. and Rogers, M.K. Spacing effects in picture memory. *Memory and Cognition*, 1973, 1, 430-434.

Howe, M.J.A. Recognition memory for photographs in homogeneous sequences. *Perceptual and Motor Skills*, 1967, 24, 1181-1182.

Hyde, T.S. and Jenkins, J.J. The differential effects of incidental tasks on the organization of recall of a list of highly associated words. *Journal of Experimental Psychology*, 1969, 82, 472-481.

Kaplan, S., Kaplan, R. and Sampson, J.R. Encoding and arousal factors in free recall of verbal and visual material. *Psychonomic Science*, 1968, 12, 73-74.

Kintsch, W. Models for free recall and recognition. In D.A. Norman (Ed.), *Models of human memory*. New York: Academic Press, 1970.

Kintsch, W. Notes on the structure of semantic memory. In E. Tulving and W. Donaldson (Eds.), *Organization of Memory*. New York: Academic Press, 1972, pp.247-308.

Kolers, P.A. Some modes of representation. In P. Pliner, L. Krames and T.M. Alloway (Eds.), *Communication and affect: language and thought*. New York: Academic Press, 1973.

Kosslyn, S.M. Scanning visual images: some structural implications. *Perception and Psychophysics*, 1973, 14, 90-94.

Lantz, D. and Stefflre, V. Language and cognition revisited. *Journal of Abnormal and Social Psychology*, 1964, 69, 472-481.

Laughery, K.R., Alexander, J.E. and Lane, A.B. Recognition of human faces: effects of target exposure time, target position, pose position and type of photograph. *Journal of Applied Psychology*, 1971, 55, 477-483.

Lieberman, L.R. and Culpepper, J.T. Words versus objects: comparison of free verbal recall. *Psychological Reports*, 1965, 17, 983-988.

Locher, P.J. and Nodine, C.F. The role of scanpaths in the recognition of random shapes. *Perception and Psychophysics*, 1974, 15, 308-314.

Lockhart, R.S. and Murdock, B.B., Jr. Memory and the theory of signal detection. *Psychological Bulletin*, 1970, 74, 100-109.

Lockhart, R.S., Craik, F.I.M. and Jacoby, L.L. Depth of processing in recognition and recall: some aspects of a general memory system. In John Brown (Ed.), *Recognition and recall*. London: Wiley, in press.

Lutz, W.J. and Scheirer, C.J. Coding processes for pictures and words. *Journal of Verbal Learning and Verbal Behavior*, 1974, 13, 316-320.

McCall, R.B. and Appelbaum, M.I. Bias in the analysis of repeated-measures designs: some alternative approaches. *Child Development*, 1973, 44, 401-405.

McCormack, P.D. Recognition memory: how complex a retrieval system? *Canadian Journal of Psychology*, 1972, 26, 19-41.

McLaughlin, B. "Intentional" and "incidental" learning in human subjects: the role of instructions to learn and motivation. *Psychological Bulletin*, 1965, 63, 359-376.

McNicol, D. *Primer of signal detection theory*. Sydney: Australasian Publishing Co., 1972.

McNulty, J.A. A partial learning model of recognition memory. *Canadian Journal of Psychology*, 1966, 20, 302-315.

Madigan, S. and Rouse, M. Picture memory and visual generation processes. *American Journal of Psychology*, 1974, 87, 151-158.

Malpass, R.S., Lavigueur, H. and Weldon, D.E. Verbal and visual training in face recognition. *Perception and Psychophysics*, 1973, 14, 285-292.

Mandler, G. Memory storage and retrieval: some limits on the reach of attention and consciousness. In P.M.A. Rabbitt and S. Dornic (Eds.), *Attention and performance V*. London: Academic Press, 1975, pp.499-516.

Mandler, J.M. and Day, J. Memory for orientation of forms as a function of their meaningfulness and complexity. *Journal of Experimental Child Psychology*, in press.

Mandler, J.M. and Stein, N.L. Recall and recognition of pictures by children as a function of organization and distractor similarity. *Journal of Experimental Psychology*, 1974, 102, 657-669.

Marken, R.S. and Sandusky, A.J. Stimulus probability and sequential effect in recognition memory. *Bulletin of the Psychonomic Society*, 1974, 4, 49-51.

Mechanic, A. The distribution of recalled items in simultaneous intentional and incidental learning. *Journal of Experimental Psychology*, 1962, 63, 593-600.

Miller, G.A. and Selfridge, J.A. Verbal context and the recall of meaningful material. *American Journal of Psychology*, 1950, 63, 176-185.

Milner, B. Visual recognition and recall after right temporal-lobe excision in man. *Neuropsychologia*, 1968, 6, 191-210.

Mooney, C. Recognition of novel visual configurations with and without eye movements. *Journal of Experimental Psychology*, 1958, 56, 133-138.

Mooney, C. Recognition of ambiguous and unambiguous visual configurations with short and longer exposures. *British Journal of Psychology*, 1960, 51, 119-125.

Morton, J. A functional model for memory. In D.A. Norman (Ed.), *Models of human memory*. New York: Academic Press, 1970.

Murphy, R.J. Recognition memory for sequentially presented pictorial and verbal spatial information. *Journal of Experimental Psychology*, 1973, 100, 327-334.

Moray, N. *Attention: selective processes in vision and hearing*. London: Hutchinson Educational Ltd., 1969.

Neisser, U. *Cognitive psychology*. New York: Appleton-Century-Crofts, 1967.

Nelson, K.E. Memory development in children: evidence from nonverbal tasks. *Psychonomic Science*, 1971, 25, 346-348.

Nelson, T.O., Metzler, J. and Reed, D.A. Role of details in the long-term recognition of pictures and verbal descriptions. *Journal of Experimental Psychology*, 1974, 102, 184-186.

Nickerson, R.S. Short-term memory for complex meaningful visual configurations: a demonstration of capacity. *Canadian Journal of Psychology*, 1965, 19, 155-160.

Nickerson, R.S. A note on long-term recognition memory for pictorial material. *Psychonomic Science*, 1968, 11, 58.

Nickerson, R.S. Auditory codability and the short-term retention of visual information. *Journal of Experimental Psychology*, 1972, 95, 429-436.

Noble, C.E. An analysis of meaning. *Psychological Review*, 1952, 59, 421-430.

Norman, D.A. (Ed.). *Models of human memory*. New York: Academic Press, 1970.

Norman, D.A. and Rumelhart, D.E. A system for perception and memory. In D.A. Norman (Ed.), *Models of human memory*. New York: Academic Press, 1970, pp.19-64.

Paivio, A. *Imagery and verbal processes.* New York: Holt, Rinehart and Winston, 1971.

Paivio, A. Spacing of repetitions in the incidental and intentional free recall of pictures and words. *Journal of Verbal Learning and Verbal Behavior*, 1974, 13, 497-511.

Paivio, A. and Csapo, K. Concrete-image and verbal memory codes. *Journal of Experimental Psychology*, 1969, 80, 279-285.

Paivio, A., Rogers, T.B. and Smythe, P.C. Why are pictures easier to recall than words? *Psychonomic Science*, 1968, 11, 137-138.

Phillips, W.A. and Baddeley, A.D. Reaction time and short-term visual memory. *Psychonomic Science*, 1971, 22, 73-74.

Posner, M.I. Abstraction and the process of recognition. In G.H. Bower and J.T. Spence (Eds.), *Psychology of learning and motivation*, Vol. 3. New York: Academic Press, 1969.

Posner, M.I. and Konick, A.F. Short-term retention of visual and kinaesthetic information. *Organizational Behavior and Human Performance*, 1966, 1, 71-85.

Posner, M.I., Boies, S.J., Eichelman, W.H. and Taylor, R.L. Retention of visual and name codes of single letters. *Journal of Experimental Psychology Monograph*, 1969, 79 (1, Pt.2).

Postman, L. Acquisition and retention of consistent associative responses. *Journal of Experimental Psychology*, 1964, 67, 183-190.

Postman, L. and Adams, P.A. Studies in incidental learning: VI. Intraserial interference. *Journal of Experimental Psychology*, 1957, 54, 153-167.

Postman, L., Adams, P.A. and Phillips, L.W. Studies in incidental learning: II. The effects of association value and of the method of testing. *Journal of Experimental Psychology*, 1955, 49, 1-10.

Potter, M.C. and Levy, E.I. Recognition memory for a rapid sequence of pictures. *Journal of Experimental Psychology*, 1969, 81, 10-15.

Prentice, W.C.H. Visual recognition of verbally labelled figures. *American Journal of Psychology*, 1954, 67, 315-320.

Price, R.H. Effects of complexity, association value, and learning on the recognition of random shapes. *Canadian Journal of Psychology*, 1968, 22, 388-398.

Price, R.H. and Slive, A.B. Verbal processes in shape recognition. *Journal of Experimental Psychology*, 1970, 83, 373-379. (a)

Price, R.H. and Slive, A.B. Association value and label relevance in shape recognition. *Perceptual and Motor Skills*, 1970, 31, 43-46. (b)

Ranken, H.B. Effects of name learning on serial learning, position learning, and recognition learning with random shapes. *Psychological Reports*, 1963, 13, 663-678.

Reed, S.K. *Psychological processes in pattern recognition.* New York: Academic Press, 1973.

Reed, S.K. Structural descriptions and the limitations of visual images. *Memory and Cognition*, 1974, 2, 329-336.

Reed, S.K. and Johnsen, J.A. Detection of parts in patterns and images. *Memory and Cognition*, 1975, 3, 569-575.

Rock, I. and Engelstein, P. A study of memory for visual form. *American Journal of Psychology*, 1959, 72, 221-229.

Rosenberg, S., Schiller, W. and Smith, J.A. Semantic factors in intentional and incidental sentence recall. *Bulletin of the Psychonomic Society*, 1974, 4, 19-21.

Rumelhart, D.E., Lindsay, P.H. and Norman, D.A. A process model for long-term memory. In E. Tulving and W. Donaldson (Eds.), *Organization of memory.* New York: Academic Press, 1972, pp.197-246.

Sampson, J.R. Free recall of verbal and non-verbal stimuli. *Quarterly Journal of Experimental Psychology*, 1970, 22, 215-221.

Sang, D. and Ross, J. The verbal loop hypothesis (VLH): A within-subject study with a perceptual recognition task. *Psychonomic Science*, 1970, 18, 345-347.

Santa, J.L. and Ranken, H.B. Effects of verbal coding on recognition memory. *Journal of Experimental Psychology*, 1972, 93, 268-278.

Santa, J.L. Verbal coding and redintegrative memory for shapes. *Journal of Experimental Psychology: Human Learning and Memory*, 1975, 104, 286-294.

Scapinello, K.F. and Yarmey, D. The role of familiarity and orientation in immediate and delayed recognition of pictorial stimuli. *Psychonomic Science*, 1970, 21, 329-331.

Scheffé, H.A. A method for judging all possible contrasts in the analysis of variance. *Biometrika*, 1953, 40, 87-104.

Schulman, A.I. Word length and rarity in recognition memory. *Psychonomic Science*, 1967, 9, 211-212.

Schulman, A.I. Recognition memory for targets from scanned word lists. *British Journal of Psychology*, 1971, 62, 335-346.

Shaffer, W.O. and Shiffrin, R.M. Rehearsal and storage of visual information. *Journal of Experimental Psychology*, 1972, 92, 292-296.

Sheehan, P.W. Visual imagery and the organizational properties of perceived stimuli. *British Journal of Psychology*, 1967, 58, 247-252.

Sheehan, P.W. A functional analysis of the role of visual imagery in unexpected recall. In P.W. Sheehan (Ed.), *The function and nature of imagery*. New York: Academic Press, 1972.

Sheehan, P.W. Stimulus imagery effect and the role of imagery in incidental learning. *Australian Journal of Psychology*, 1973, 25, 93-102.

Shepard, R.N. and Teghtsoonian, M. Retention of information under conditions approaching a steady state. *Journal of Experimental Psychology*, 1961, 62, 302-309.

Siegel, A.W., Babich, J.M. and Kirasic, K.C. Visual recognition memory in reflective and impulsive children. *Memory and Cognition*, 1974, 2, 379-384.

Smith, E.E. and Larson, D.E. The verbal loop hypothesis and the effects of similarity on recognition and communication in adults and children. *Journal of Verbal Learning and Verbal Behavior*, 1970, 9, 237-242.

Smith, E.E. and Nielsen, G.D. Representations and retrieval processes in short-term memory: recognition and recall of faces. *Journal of Experimental Psychology*, 1970, 85, 397-405.

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

Standing, L. Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology*, 1973, 25, 207-222.

Standing, L, Conezio, J. and Haber, R.N. Perception and memory for pictures: single-trial learning of 2500 visual stimuli. *Psychonomic Science*, 1970, 19, 73-74.

Stein, N.L. and Mandler, J.M. Children's recognition of reversals of geometric figures. *Child Development*, 1974, 45, 604-615.

Stein, N.L. and Mandler, J.M. Development of detection and recognition of orientation of geometric and real figures. *Child Development*, 1975, 46, 379-388.

Street, R.F. *A Gestalt completion test*. New York: Teachers College, Columbia University, 1931.

Sutherland, N.S. Outlines of a theory of visual pattern recognition in animals and man. *Proceedings of the Royal Society*, 1968, 171, 297-317.

Ternes, W. and Yuille, J.C. Words and pictures in a short-term memory task. *Journal of Experimental Psychology*, 1972, 96, 78-86.

Tulving, E. Episodic and semantic memory. In E. Tulving and W. Donaldson (Eds.), *Organization of memory*. New York: Academic Press, 1972, pp.381-403.

Tulving, E. and Madigan, S.A. Memory and verbal learning. *Annual Review of Psychology*, 1970, 21, 437-484.

Tversky, B. Encoding processes in recognition and recall. *Cognitive Psychology*, 1973, 5, 275-287.

Tversky, B. Eye fixations in prediction of recognition and recall. *Memory and Cognition*, 1974, 2, 275-278.

Tversky, B. and Sherman, T. Picture memory improves with longer on time and off time. *Journal of Experimental Psychology*, 1975, 104, 114-118.

Tzavaras, A., Hecaen, H. and Le Bras, H. Troubles de la reconnaissance du visage humain et latéralisation hémisphérique lesionelle chez les sujets gauchers. *Neuropsychologia*, 1971, 9, 475-477.

Underwood, B.J. *Experimental psychology*. New York: Appleton-Century-Crofts, 1949.

Vanderplas, J.M. and Garvin, E.A. The association value of random shapes. *Journal of Experimental Psychology*, 1959, 57, 147-157.

von Wright, J.M. Relation between verbal recall and visual recognition of the same stimuli in young children. *Journal of Experimental Child Psychology*, 1973, 15, 481-487.

Warrington, E. and James, M. An experimental investigation of facial recognition in patients with localized cerebral lesions. *Cortex*, 1967, 3, 317-326.

Watkins, M.J. When is recall spectacularly higher than recognition? *Journal of Experimental Psychology*, 1974, 102, 161-163.

Waugh, N.C. and Norman, D.A. Primary memory. *Psychological Review*, 1965, 72, 89-104.

Winer, B.J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1971 (Second Edition).

Wiseman, S. and Neisser, U. Perceptual organization as a determinant of visual recognition memory. *American Journal of Psychology*, 1974, 87, 675-681.

Wulf, F. Beiträge zur Psychologie der Gestalt; vi. Ueber die Veränderung von Vorstellungen (Gedächtnis und Gestalt), *Psychologische Forschung*, 1922, 1, 333-373.

Wyant, S., Banks, W.P., Berger, D. and Wright, P.W. Verbal and pictorial similarity in recognition of pictures. *Perception and Psychophysics*, 1972, 12, 151-153.

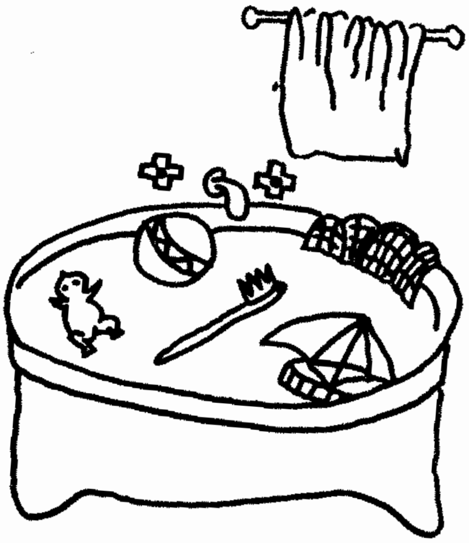
Yarmey, A.D. Recognition memory for familiar "public" faces: effects of orientation and delay. *Psychonomic Science*, 1971, 24, 286-288.

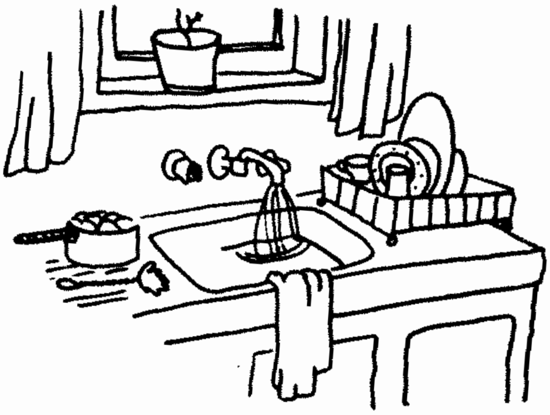
Yin, R.K. Looking at upside-down faces. *Journal of Experimental Psychology*, 1969, 81, 141-145.

APPENDIX 1

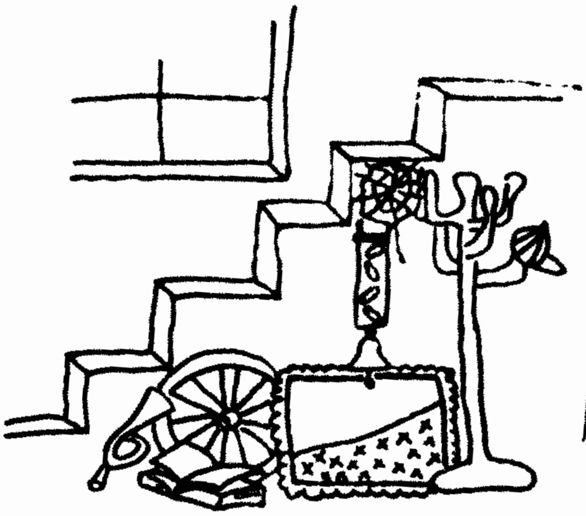
Practice and test stimuli used in Experiment 1.
A complete distractor set is shown for Prototype 3.

Practice Items



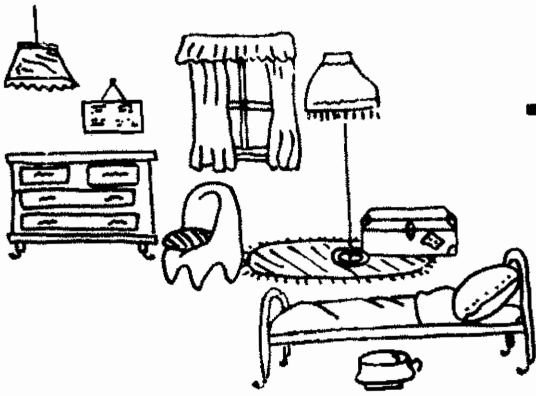


Test Items



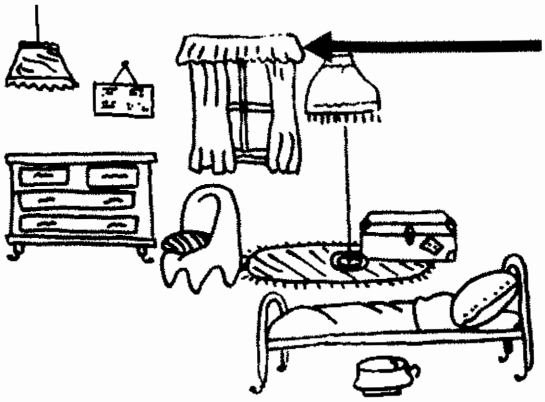
1.

4.

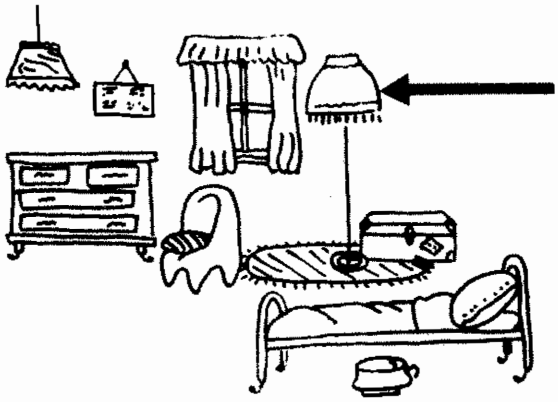


Prototype 3.

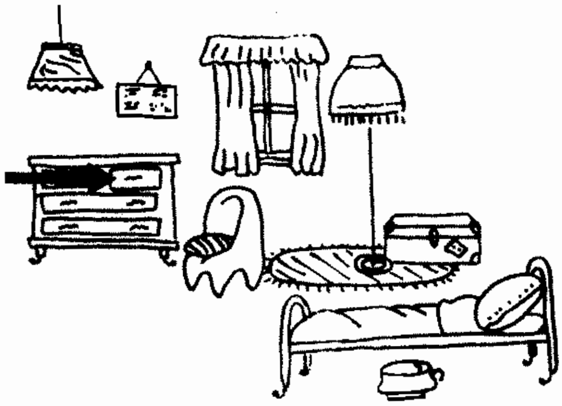
V1.



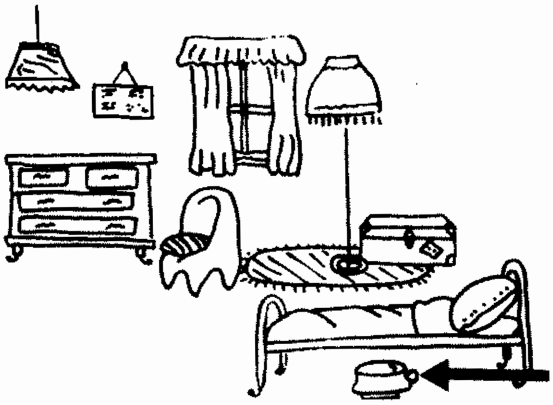
V2.



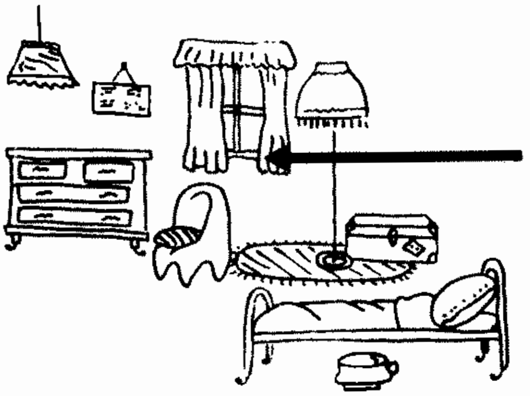
V3.



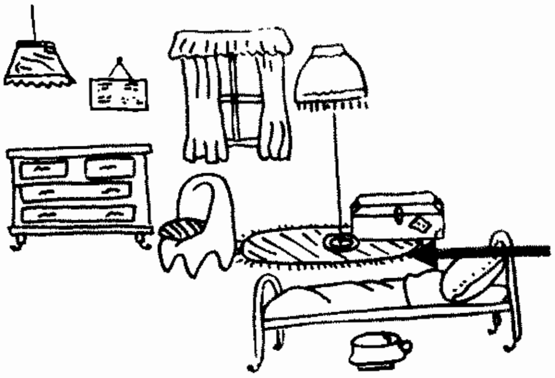
V4.



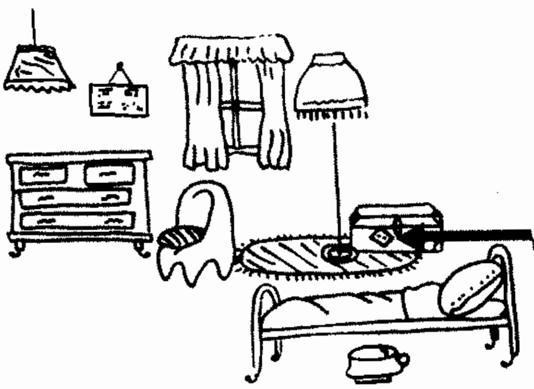
V5.



V6.

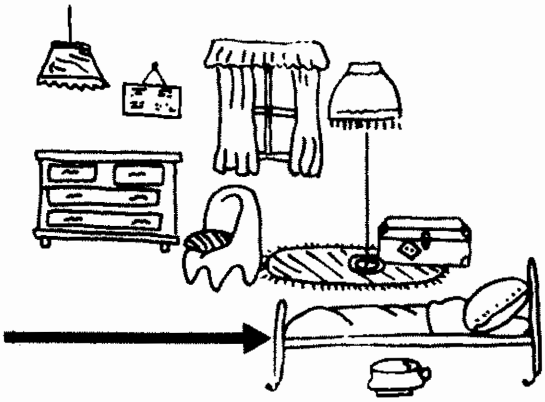


V7.

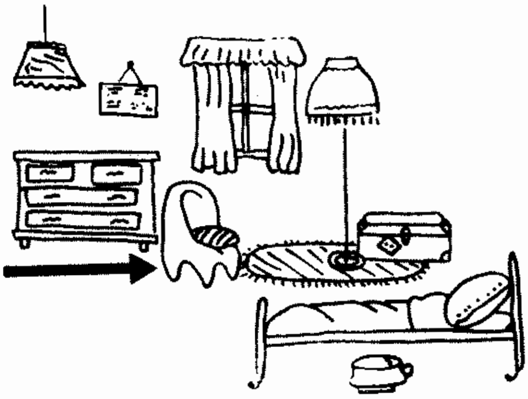


V8.

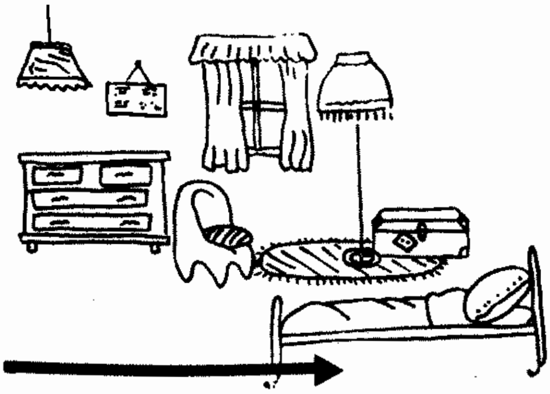
V9.



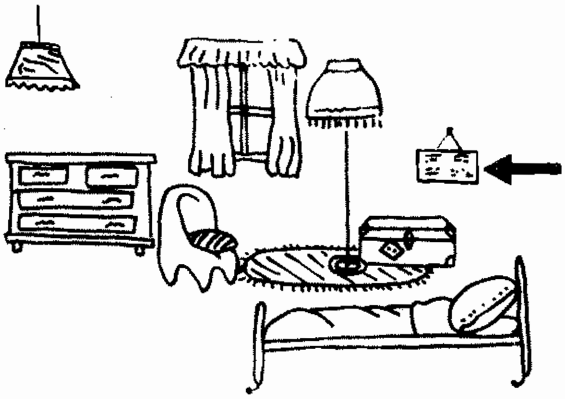
V10.



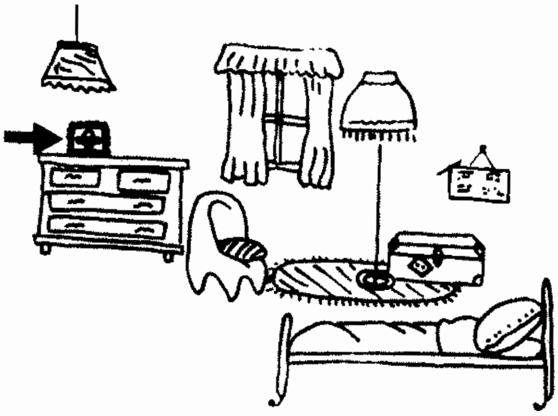
V11.



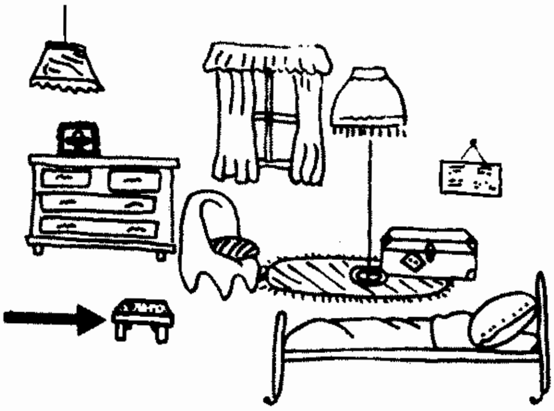
V12.



V13.



V14.



V15.

APPENDIX 2

The questionnaire given to subjects in Experiment 2.

QUESTION SHEET

This experiment was used to study how our picture memory works. Different people see the task in different ways. What did you think the purpose of the experiment was?

What exactly were you doing when you were looking at the pictures originally?

Did you find anything difficult?

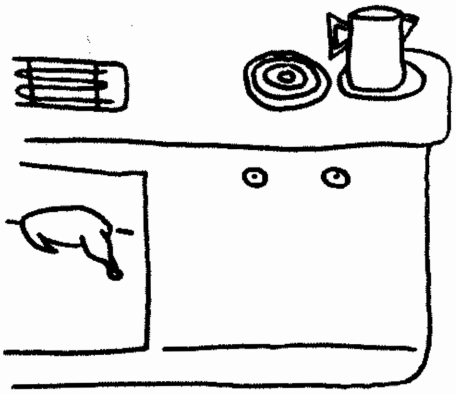
What did you think when you were asked about your memory for the pictures?

Were you purposely trying to remember the pictures when you saw them originally?

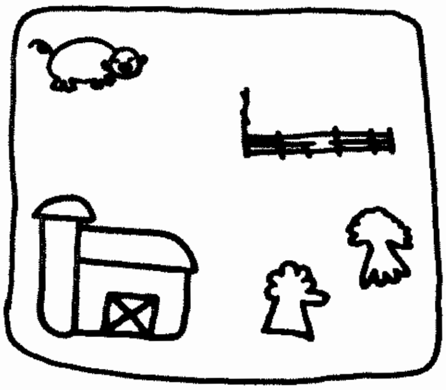
Did you do anything in particular in the time *between* presentations of the pictures (when first shown them)?

APPENDIX 3

Materials and tables associated with Experiment 3.

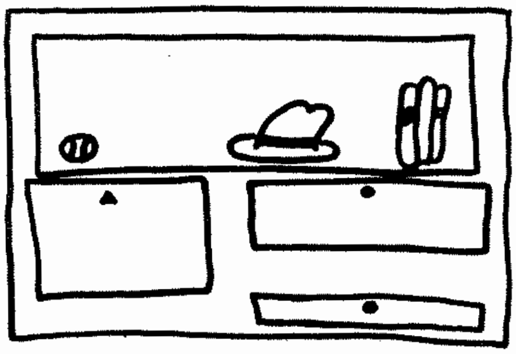


Practice

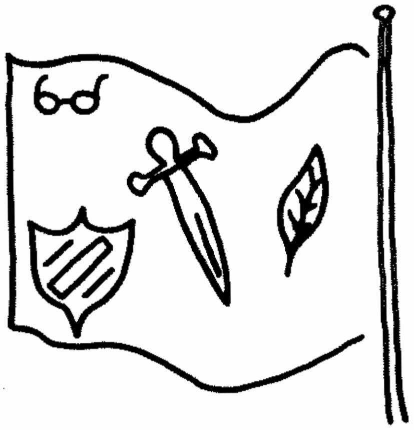


1.

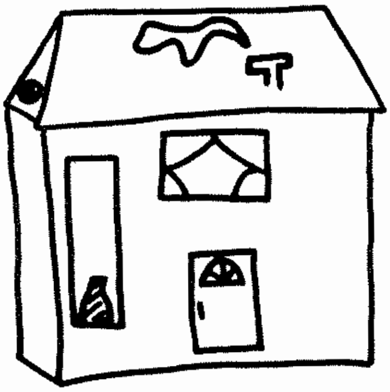
2.



4.

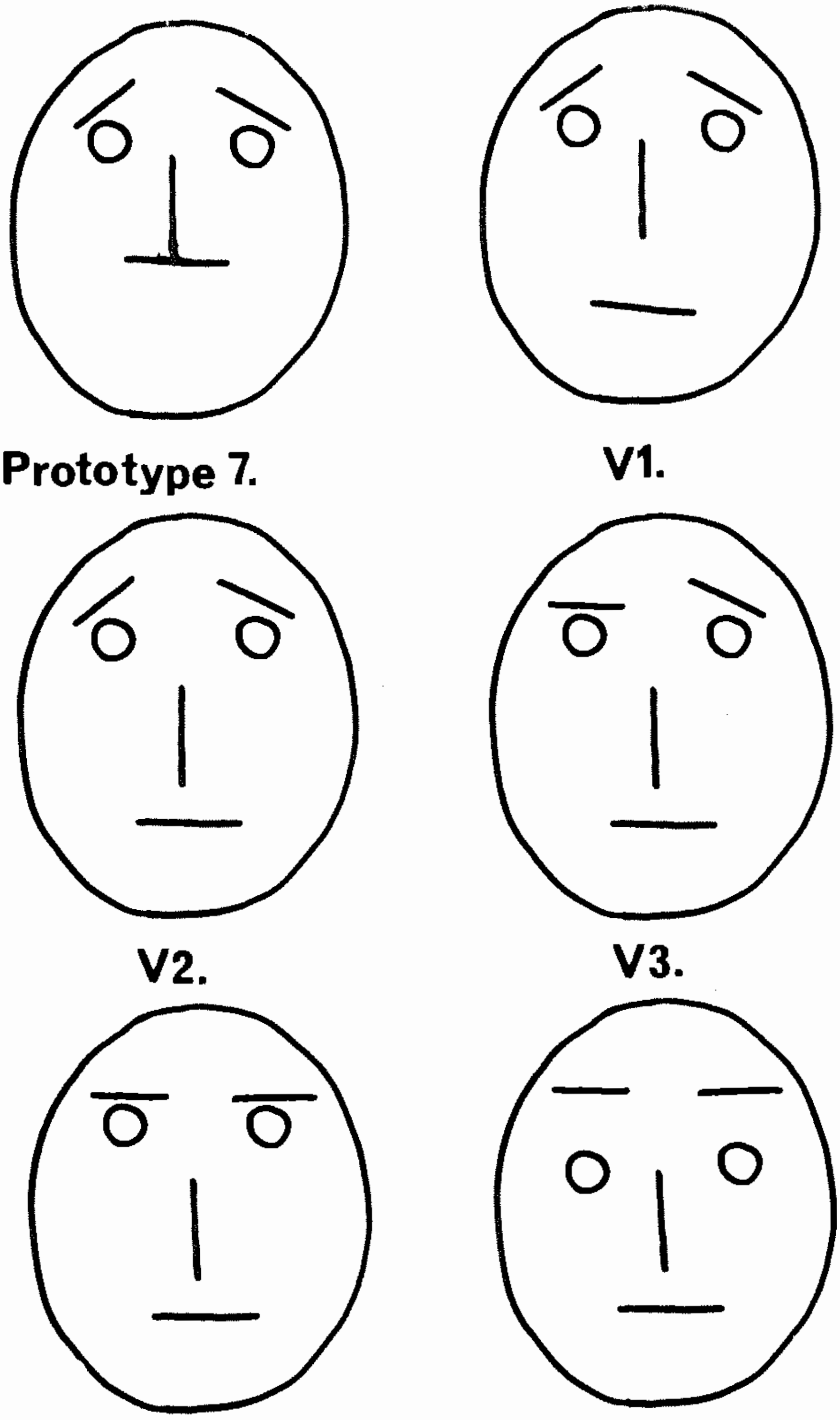


5.



6.

The practice stimulus and Pictures 1 to 6 in Experiment 3.



Prototype 7.

V1.

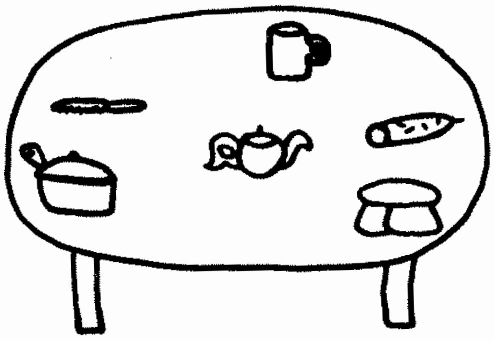
V2.

V3.

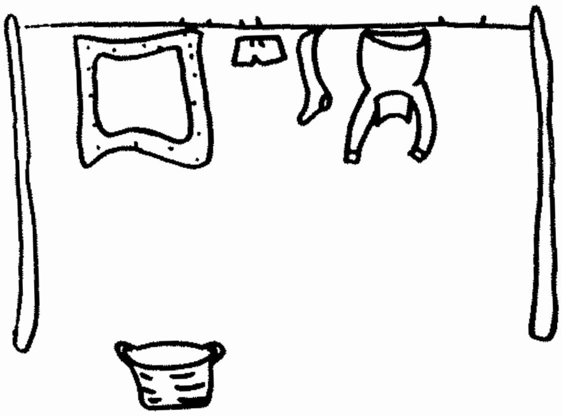
V4.

V5.

Picture 7 and its five distractor variations.

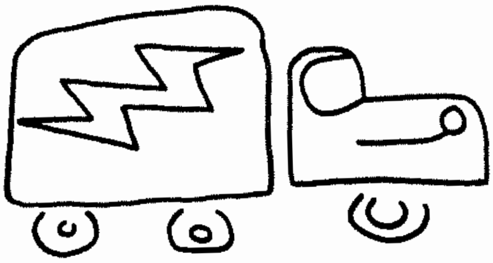


8.



9.

10.



11.

Pictures 8 to 11 in Experiment 3.

Table A1

Means and standard deviations of exclusion set scores for each picture, in each instructional group in Experiment 3.

Picture		Group		
		Incidental	Control	Intentional
1.	$\bar{X}$	2.8667	2.1333	3.0667
	S.D.	1.8465	1.1872	2.0517
2.	$\bar{X}$	2.3333	1.2000	3.5333
	S.D.	1.9881	1.6125	1.5976
3.	$\bar{X}$	1.9333	2.5333	3.5333
	S.D.	1.8695	1.8848	1.1255
4.	$\bar{X}$	3.0000	2.2667	4.0000
	S.D.	1.5584	1.5338	1.2536
5.	$\bar{X}$	3.0000	2.7333	3.4667
	S.D.	1.6903	1.8310	1.8465
6.	$\bar{X}$	2.6000	2.5333	2.2000
	S.D.	1.4541	1.9952	1.7808
7.	$\bar{X}$	2.7333	3.4000	3.8667
	S.D.	1.2799	1.1212	1.4573
8.	$\bar{X}$	1.4667	1.0000	2.4000
	S.D.	1.5523	1.1952	1.9928
9.	$\bar{X}$	3.4667	4.0000	4.0000
	S.D.	1.4573	1.5584	1.6036
10.	$\bar{X}$	2.0000	1.6667	3.0667
	S.D.	1.5119	1.8387	1.8695
11.	$\bar{X}$	2.0667	1.3333	3.1333
	S.D.	1.9809	1.9149	1.9591
Total	$\bar{X}$	2.4970	2.2545	3.2970
	S.D.	0.6608	0.7723	0.9369

Table A2

Means and standard deviations of correct rejection scores for all pictures in each instructional group in Experiment 3.

Picture		Group		
		Incidental	Control	Intentional
1.	$\bar{X}$	.8678	.8556	.9033
	S.D.	.1386	.1849	.2621
2.	$\bar{X}$	.7144	.6922	.7933
	S.D.	.3887	.3320	.3385
3.	$\bar{X}$	.7878	.8589	.9589
	S.D.	.3341	.1603	.0866
4.	$\bar{X}$	.9233	.8233	.9133
	S.D.	.1175	.2588	.1153
5.	$\bar{X}$	.9067	.8311	.9622
	S.D.	.1233	.2579	.0787
6.	$\bar{X}$	.7511	.8911	.8767
	S.D.	.3362	.1435	.1803
7.	$\bar{X}$	.7889	.9022	.9256
	S.D.	.1010	.1100	.1109
8.	$\bar{X}$	.8056	.7978	.9067
	S.D.	.1985	.2592	.1233
9.	$\bar{X}$	.8678	.9178	.9289
	S.D.	.1349	.1068	.0906
10.	$\bar{X}$	.8356	.7756	.8178
	S.D.	.2715	.3471	.3477
11.	$\bar{X}$	.7067	.8111	.7889
	S.D.	.3830	.3557	.2970
Total	$\bar{X}$	.8141	.8324	.8887

Table A3

Summary of the analysis of variance of
correct rejection scores in Experiment 3.

Source of Variation	SS	df	MS	F	p
Between Subjects	3.0976	44			
Instructions	0.4977	2	0.2489	4.0210	< .05
Subjects within Groups	2.5999	42	0.0619		
Within Subjects	25.4671	450			
Pictures	1.3580	10	0.1358	2.4293	< .01
Instructions × Pictures	0.6481	20	0.0324	0.5796	NS
Pictures × Subjects within Groups	23.4610	420	0.0559		

Table A4

Means and standard deviations of hit rate scores for all pictures in each instructional group in Experiment 3.

Picture		Group		
		Incidental	Control	Intentional
1.	$\bar{X}$	.1844	.2067	.5156
	S.D.	.2780	.2755	.3867
2.	$\bar{X}$	.1511	.1244	.2167
	S.D.	.1510	.1581	.1910
3.	$\bar{X}$	.3167	.2056	.4500
	S.D.	.3961	.2884	.3390
4.	$\bar{X}$	.3744	.2244	.3889
	S.D.	.3752	.2912	.3867
5.	$\bar{X}$	.3556	.1667	.5578
	S.D.	.3500	.1967	.4093
6.	$\bar{X}$	.1667	.2611	.2756
	S.D.	.1863	.3002	.3415
7.	$\bar{X}$	.0444	.3222	.5111
	S.D.	.1175	.3908	.4475
8.	$\bar{X}$	.1578	.1211	.2689
	S.D.	.2135	.1425	.3300
9.	$\bar{X}$	.2722	.3944	.4222
	S.D.	.3597	.4184	.4173
10.	$\bar{X}$	.2211	.2600	.3889
	S.D.	.2700	.2795	.3129
11.	$\bar{X}$	.2444	.3333	.3889
	S.D.	.2877	.3197	.4211
Total	$\bar{X}$	.2278	.2382	.3986
	S.D.	.0825	.0848	.1350

Table A5

Means and standard deviations of circles counted in each orienting task group, for each picture, in Experiment 3.

Picture		Group	
		Incidental	Control
1.	$\bar{X}$	4.5333	2.1333
	S.D.	2.9244	1.9223
2.	$\bar{X}$	3.3333	2.4667
	S.D.	2.5820	1.3020
3.	$\bar{X}$	4.5333	3.8000
	S.D.	1.8465	1.4243
4.	$\bar{X}$	4.6667	3.4000
	S.D.	1.6330	0.7368
5.	$\bar{X}$	3.8667	2.6667
	S.D.	1.4075	0.8165
6.	$\bar{X}$	1.6667	1.0000
	S.D.	1.3452	0.0000
7.	$\bar{X}$	3.0000	3.0000
	S.D.	0.0000	0.0000
8.	$\bar{X}$	5.9333	4.6000
	S.D.	2.6040	1.8439
9.	$\bar{X}$	3.8667	1.6000
	S.D.	3.3566	0.9103
10.	$\bar{X}$	3.7333	7.3333
	S.D.	5.4963	1.4475
11.	$\bar{X}$	2.6667	5.7333
	S.D.	5.3940	2.4339
Total	$\bar{X}$	4.2242	3.0000
	S.D.	1.6854	0.9706

Table A6

Means (and standard deviations) of confidence ratings for each prototype and its five variations in each instructional group in Experiment 3.

Picture/Group	Prototype	Variation 1	Variation 2	Variation 3	Variation 4	Variation 5
1. Incidental	3.867 (1.506)	3.933 (1.387)	3.467 (1.302)	3.933 (1.163)	4.467 (1.302)	5.467 (1.126)
Control	3.400 (0.828)	3.467 (0.834)	3.200 (1.082)	3.400 (1.121)	3.600 (1.986)	5.600 (.6325)
Intentional	2.333 (1.175)	3.867 (1.552)	4.200 (1.568)	4.467 (1.356)	4.867 (1.187)	5.133 (1.302)
2. Incidental	3.533 (1.408)	2.733 (0.961)	3.200 (1.146)	3.200 (1.082)	3.733 (1.335)	4.000 (1.604)
Control	3.600 (1.121)	3.133 (0.991)	3.067 (1.033)	3.933 (1.100)	3.200 (0.862)	3.733 (1.223)
Intentional	3.267 (1.387)	2.667 (1.047)	3.867 (1.356)	4.133 (1.302)	3.733 (1.487)	4.400 (1.454)
3. Incidental	3.800 (1.373)	4.067 (1.280)	3.733 (1.438)	4.667 (1.235)	4.600 (1.298)	4.267 (1.387)
Control	3.400 (1.298)	3.333 (1.175)	4.000 (1.000)	3.800 (1.082)	4.133 (1.060)	4.267 (1.163)
Intentional	2.467 (1.506)	3.267 (1.751)	4.333 (1.676)	5.133 (1.302)	5.533 (0.743)	5.667 (0.617)
4. Incidental	3.133 (1.457)	4.000 (1.414)	4.867 (1.407)	4.600 (1.549)	4.867 (1.506)	5.600 (0.910)
Control	3.400 (1.502)	3.467 (1.356)	3.667 (1.447)	3.933 (1.280)	4.200 (1.320)	5.200 (1.082)
Intentional	2.867 (1.302)	3.333 (1.447)	5.333 (0.900)	5.267 (1.033)	5.267 (1.280)	5.733 (0.799)
5. Incidental	3.600 (1.595)	3.733 (1.534)	4.400 (1.183)	3.867 (1.506)	5.467 (0.915)	5.267 (1.163)
Control	3.467 (1.302)	3.067 (1.163)	3.933 (1.223)	3.933 (0.884)	5.267 (0.961)	5.400 (1.055)
Intentional	2.467 (1.356)	4.067 (1.624)	4.133 (1.356)	4.533 (1.187)	5.667 (0.724)	5.400 (0.985)
6. Incidental	3.733 (1.387)	3.133 (1.506)	3.333 (1.543)	4.800 (1.207)	4.400 (1.502)	4.933 (1.387)
Control	3.667 (1.345)	4.067 (1.163)	4.267 (1.223)	4.267 (1.280)	4.333 (0.900)	4.667 (1.543)
Intentional	3.200 (1.473)	3.200 (1.082)	3.533 (1.506)	3.800 (1.424)	4.067 (1.387)	5.467 (1.187)

Table A6 (cont'd)

Picture/Group	Prototype	Variation 1	Variation 2	Variation 3	Variation 4	Variation 5
7. Incidental	5.000 (1.464)	3.400 (1.805)	3.333 (1.877)	5.533 (0.834)	4.733 (1.751)	4.933 (1.580)
Control	3.467 (1.807)	3.667 (1.291)	3.800 (1.265)	5.533 (0.743)	5.000 (1.414)	5.067 (1.033)
Intentional	3.133 (1.685)	4.267 (1.710)	3.667 (1.589)	5.800 (0.560)	5.467 (1.126)	5.800 (0.560)
8. Incidental	3.867 (1.356)	3.467 (1.457)	3.733 (1.624)	4.000 (1.414)	4.000 (1.414)	4.800 (1.373)
Control	3.600 (1.055)	3.467 (1.060)	3.333 (1.113)	3.600 (1.055)	3.267 (1.033)	4.733 (1.033)
Intentional	3.067 (1.163)	3.667 (1.589)	3.867 (1.552)	3.933 (1.223)	3.800 (1.265)	5.267 (0.961)
9. Incidental	3.933 (1.486)	3.733 (1.486)	4.533 (1.408)	4.600 (1.502)	4.600 (1.352)	5.533 (0.743)
Control	3.333 (1.448)	3.933 (1.486)	4.533 (1.302)	4.733 (1.223)	5.000 (1.000)	5.467 (0.640)
Intentional	3.200 (1.568)	4.067 (1.668)	5.200 (1.265)	5.333 (0.976)	5.533 (0.743)	5.733 (0.593)
10. Incidental	3.533 (1.767)	3.600 (1.352)	3.867 (1.597)	4.333 (1.397)	4.200 (1.320)	4.800 (1.373)
Control	3.200 (1.082)	3.267 (1.100)	3.600 (0.910)	3.933 (1.580)	4.000 (1.309)	4.733 (1.335)
Intentional	2.800 (1.473)	3.267 (1.438)	3.800 (1.146)	4.333 (1.235)	4.467 (1.302)	5.133 (1.246)
11. Incidental	3.733 (1.223)	3.867 (1.356)	3.800 (1.424)	3.867 (1.187)	4.133 (1.457)	4.067 (1.624)
Control	3.200 (1.146)	3.400 (1.352)	3.533 (1.187)	3.867 (1.060)	3.800 (1.265)	4.267 (0.961)
Intentional	3.400 (1.549)	3.533 (1.356)	3.733 (1.534)	4.400 (1.404)	4.333 (1.235)	4.733 (1.163)

Table A7

Means, standard deviations and frequencies of confidence ratings for each item in the test set averaged over all pictures, for each position in the list and each instructional group in Experiment 3.

Position		Prototype	Variation 1	Variation 2	Variation 3	Variation 4	Variation 5
				INCIDENTAL GROUP			
1.	$\bar{X}$	3.0417	3.5152	3.7353	4.2414	3.9000	3.9091
	S.D.	1.3015	1.4603	1.2385	1.6830	1.7442	1.6877
	N	24	33	34	29	20	22
2.	$\bar{X}$	4.000	3.8800	3.8750	4.3030	4.3448	5.0000
	S.D.	1.4720	1.4810	1.4549	1.2621	1.3958	1.0505
	N	25	25	16	33	29	30
3.	$\bar{X}$	3.7931	3.5238	3.9333	4.3200	4.5862	5.0000
	S.D.	1.4238	1.3274	1.6386	1.4059	1.2106	1.4142
	N	29	21	30	25	29	25
4.	$\bar{X}$	3.9286	3.4800	4.0000	4.7037	4.5926	4.9000
	S.D.	1.4123	1.5578	1.5176	1.1373	1.5257	1.4473
	N	28	25	34	27	27	20
5.	$\bar{X}$	3.9615	3.9167	4.1200	4.2727	4.7083	4.8974
	S.D.	1.5616	1.5299	1.7156	1.2414	1.2329	1.3337
	N	26	24	25	22	24	39
6.	$\bar{X}$	4.1481	3.6875	3.5000	4.6250	4.6129	5.3846
	S.D.	1.5861	1.2032	1.6384	1.4390	1.4066	0.9829
	N	27	32	20	24	31	26

Table A7 (cont'd)

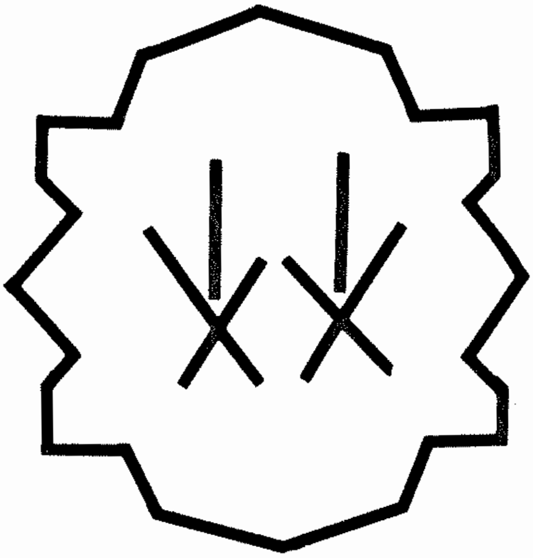
Position	Prototype	Variation 1	Variation 2	Variation 3	Variation 4	Variation 5
			CONTROL GROUP			
1. $\bar{X}$	3.2500	3.8333	3.5000	3.9259	4.1765	4.6923
S.D.	1.2443	1.4653	1.6135	1.2066	1.5098	1.0870
N	32	18	30	27	17	26
2. $\bar{X}$	3.6818	3.5909	4.4583	4.1786	4.0303	4.7619
S.D.	1.3934	1.2212	1.2504	1.2488	1.2866	1.0443
N	22	22	24	28	33	21
3. $\bar{X}$	3.3704	3.1818	4.0417	3.9565	3.9130	5.1290
S.D.	1.0795	1.0065	1.1602	1.2239	1.5049	1.0876
N	27	22	24	23	23	31
4. $\bar{X}$	3.1111	3.7500	3.5000	4.3462	4.2273	4.8750
S.D.	1.1318	1.3912	1.0000	1.0933	1.1098	1.2619
N	18	32	28	26	22	24
5. $\bar{X}$	3.6333	3.3333	3.7000	4.1364	4.4545	5.0000
S.D.	1.3257	0.9608	1.3416	1.4572	1.1750	1.0847
N	30	27	20	22	33	18
6. $\bar{X}$	3.4286	3.4138	3.1667	4.0417	4.7727	5.1000
S.D.	1.5675	1.1501	0.7614	1.4289	0.9223	1.2690
N	21	29	24	24	22	30

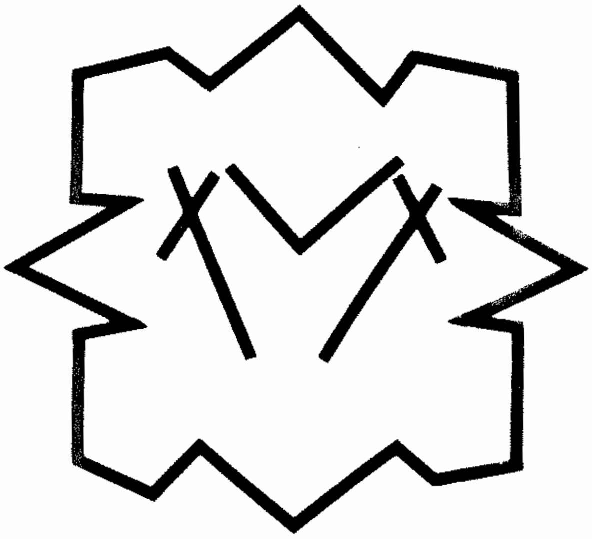
Table A7 (cont'd)

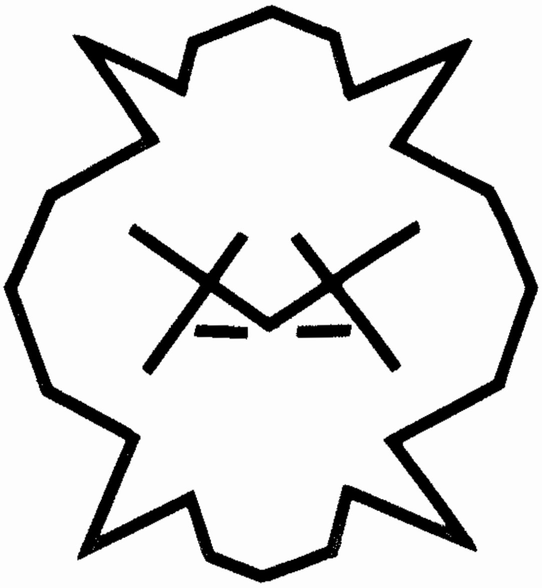
Position		Prototype	Variation 1	Variation 2	Variation 3	Variation 4	Variation 5
				INTENTIONAL GROUP			
1.	$\bar{X}$	2.8800	3.7333	3.9333	4.5909	4.7222	5.2000
	S.D.	1.4526	1.7006	1.7407	1.5632	1.5265	1.0408
	N	25	30	30	22	18	25
2.	$\bar{X}$	2.8947	3.4722	3.9615	4.6429	4.8235	5.0833
	S.D.	1.1002	1.3625	1.1483	1.4198	1.2862	1.2825
	N	19	36	26	28	17	24
3.	$\bar{X}$	3.0000	3.6071	4.0385	4.6111	4.8621	5.3500
	S.D.	1.2172	1.5477	1.4555	1.4200	1.3289	1.0400
	N	28	28	26	18	29	20
4.	$\bar{X}$	3.0476	4.3333	4.4839	4.2174	4.8966	5.5862
	S.D.	1.5645	1.6088	1.4112	1.3469	1.3187	0.9826
	N	21	18	31	23	29	29
5.	$\bar{X}$	2.7586	3.4667	4.3158	4.6129	4.9231	5.5667
	S.D.	1.6400	1.6417	1.7337	1.4760	1.2304	0.8584
	N	29	15	19	31	26	30
6.	$\bar{X}$	3.0000	3.5652	4.6667	5.1071	5.0000	5.5909
	S.D.	1.6997	1.4086	1.4951	1.1001	1.0954	0.7342
	N	28	23	18	28	31	22

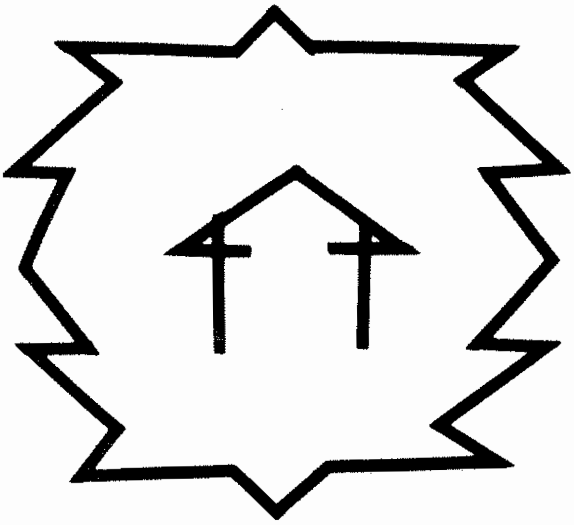
APPENDIX 4

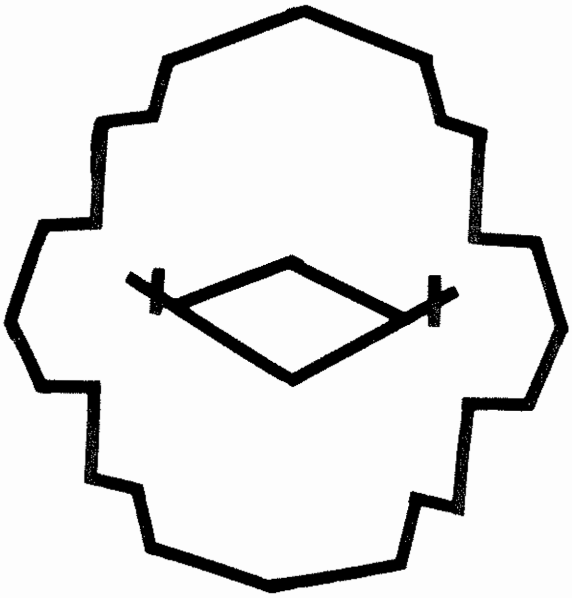
Practice and test stimuli used in Experiment 4.

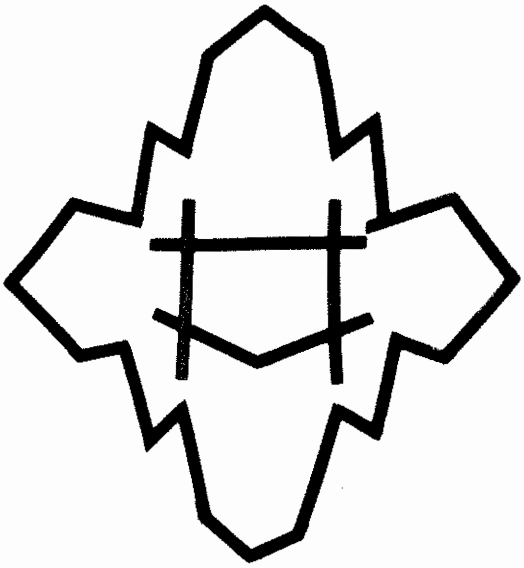


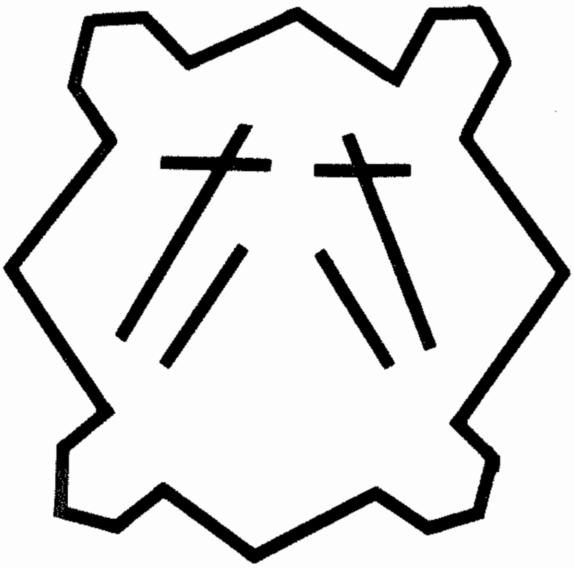






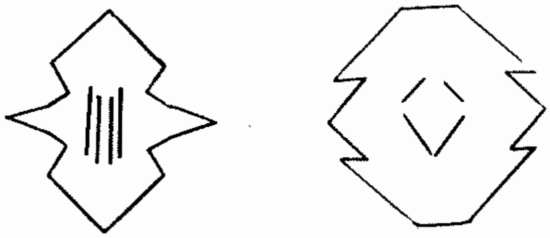






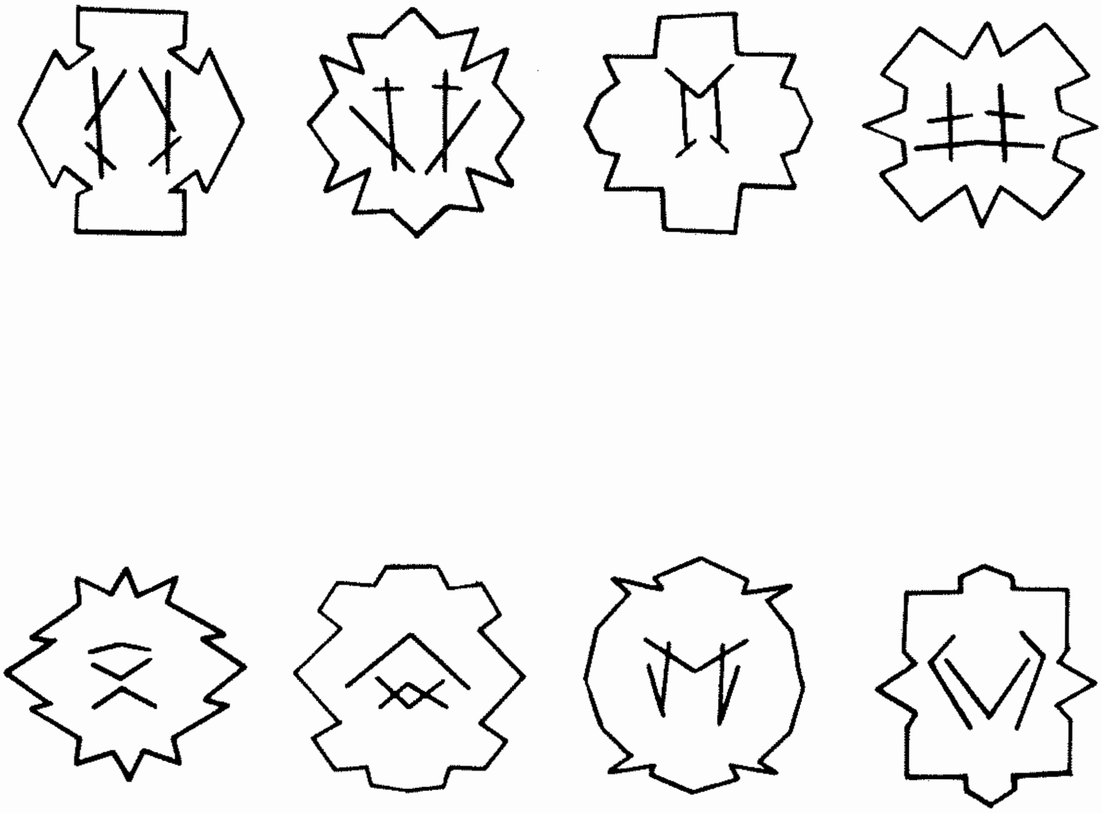
APPENDIX 5

Practice and test stimuli used in Experiment 5
and test stimuli used in both Experiments 4 and 5.

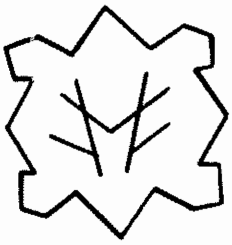
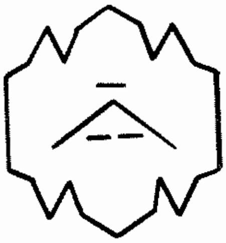
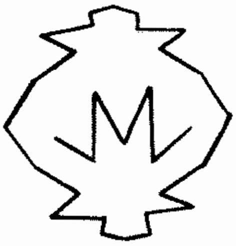
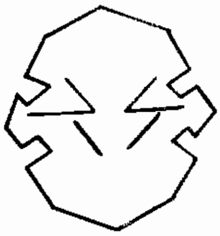
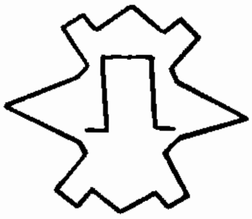
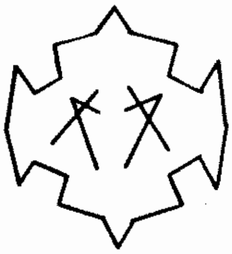
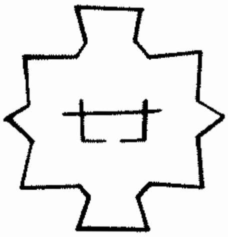
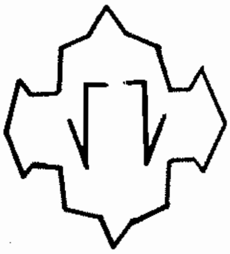
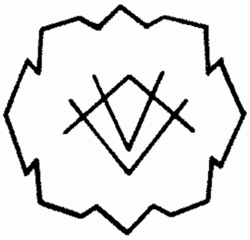
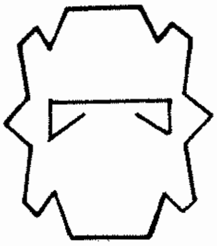
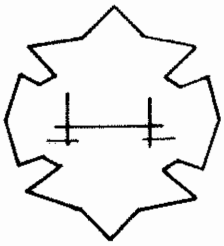
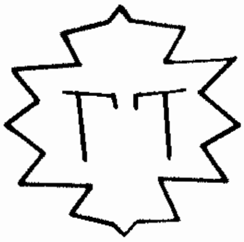
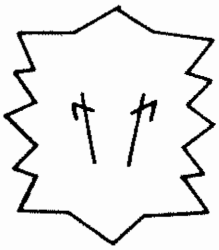
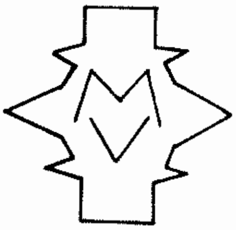


Practice (Exp. 5)

Items used in both Experiments 4 and 5



Items used in Experiment 5 only

APPENDIX 6

Practice and test stimuli used in Experiment 8.