12-1-1979

Internal and External Context Effects Upon the Types of Information Encoded From Pictures

James G. Penland

INTERNAL AND EXTERNAL CONTEXT EFFECTS UPON THE TYPES OF INFORMATION ENCODED FROM PICTURES

by
James G. Penland

Bachelor of Arts, Metropolitan State College, 1977

A Thesis

Submitted to the Graduate Faculty

of the

University of North Dakota

in partial fulfillment of the requirements

for the degree of

Master of Arts

Grand Forks, North Dakota

December
1979

Internal and External Context Effects Upon the Types of Information Encoded from Pictures

James G. Penland, M.A.

The University of North Dakota, 1979

Faculty Advisor: Dr. James Antes

Several investigators (e.g., Biederman, et al. 1974; Loftus & Bell 1975) have proposed that one's perception and subsequent recognition of pictured scenes results from the acquisition of two types of information. One is specific in nature and results from the direct inspection of object detail. The second type is more general in character and is thought to result from the processing of contextual information. Friedman (in press) has pointed out that context must be considered to encompass both internal (memorial) and external (physical) constraints, and as such to have a dual informational base. The present study was designed to assess the separate and combined effects of these internal and external sources of contextual information as they influence the amount of object detail later available to the observer.

Each of 72 (21 male) undergrads was presented with a written phrase prior to a 150 msec. exposure of a pictured scene, followed by a four alternative forced-choice recognition test. Subjects responded by selecting one of the four objects in the test as having been viewed previously, and then rated their confidence in that selection. The type of internal or memorial-based contextual information, prompted

by a written phrase describing the theme of each stimulus picture, was varied within subjects such that each subject viewed one-third of the pictures preceded by a compatible theme, one-third preceded by a neutral phrase, and one-third preceded by an incompatible theme. External context and recognition test distractors were varied between subjects. External or physical context was either present or absent, and the distractor objects were objects from dissimilar scenes, objects differing from those viewed but from similar scenes, or objects possessing the same generic name as the target but differing in some physical attribute.

Distractors from different scenes consistently resulted in the best recognition performance and the most confidence, with no differences in accuracy between the other two distractor types. Subject confidence was a more sensitive measure of distractor effects, as ratings were significantly higher when distractors were varied from the target in object rather than attribute information. Recognition accuracy and confidence were also enhanced when the pre-stimulus prompt was compatible with the stimulus picture, but only when distractor objects were from dissimilar scenes. A trend was evidenced suggesting that recognition accuracy may have been enhanced by the presence of external context when test distractors were from dissimilar scenes, whereas the absence of context may have facilitated performance when distractors differed from the target object only in physical detail.

These findings were interpreted via a consideration of the influences of the two types of information available in real world scenes and of the demands imposed on the subject by the particular task employed to study their perception and recognizability.

This Thesis submitted by James G. Penland in partial fulfillment of the requirements for the Degree of Master of Arts from the University of North Dakota is hereby approved by the Faculty Advisory Committee under whom the work has been done.

James R. Antes
(Chairman)

Mark Grabe

Paul H. Wright

This thesis meets the standards for appearance and conforms to the style and format requirements of the Graduate School of the University of North Dakota, and is hereby approved.

A. William Johnson
Dean of the Graduate School

Permission

Title Internal and External Context Effects Upon the Type of Information Encoded from Pictures

Department Psychology

Degree Master of Arts

In presenting this thesis in partial fulfillment of the requirements for a graduate degree from the University of North Dakota, I agree that the Library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by the professor who supervised my thesis work or, in his absence, by the Chairman of the Department or the Dean of the Graduate School. It is understood that any copying or publication or other use of this thesis or part thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of North Dakota in any scholarly use which may be made of any material in my thesis.

Signature James G. Penland

Date November 1, 1979

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS.................................... vi

LIST OF TABLES.................................... vii

ACKNOWLEDGMENTS.................................... viii

ABSTRACT.................................... ix

CHAPTER I. INTRODUCTION.................................... 1

A History of Process Mediation
Looking at Pictures
A Duplex Theory of Picture Perception
Context Effects in Picture Perception
Overview and Statement of the Problem

CHAPTER II. METHOD.................................... 63

Design
Subjects
Stimuli
Pictures
Prompts
Response Alternatives
Procedure
Data Analysis

CHAPTER III. RESULTS.................................... 76

Recognition Accuracy
Confidence Ratings
Weighted Confidence Scores
Accuracy, Confidence Ratings, and Weighted Confidence Scores Summarized

CHAPTER IV. DISCUSSION.................................... 95

Recognition Accuracy
Confidence Ratings
Weighted Confidence Scores
Summary and Implications

APPENDICES.. 122

APPENDIX A: INSTRUCTIONS READ TO ALL SUBJECTS.............. 123

APPENDIX B: ANSWER SHEET USED TO RECORD SUBJECTS;
RESPONSES.. 126

APPENDIX C: SUMMARY TABLES FOR POST HOC ANOVA'S
REPORTED IN TEXT.. 128

APPENDIX D: TABLE 10.. 132

REFERENCES.. 136

LIST OF ILLUSTRATIONS

Figure		Page
1.	Example of Targeted Object Presented Under the Context-Present Condition	66
2.	Example of Targeted Object Presented Under the Context-Absent Condition	67
3.	Example of Response Matrix Employed Under the Scene-Distractor Condition	70
4.	Example of Response Matrix Employed Under the Object-Distractor Condition	71
5.	Example of Response Matrix Employed Under the Attribute-Distractor Condition	72
6.	Mean Number of Correct Recognition Responses	81
7.	Mean Weighted Confidence Scores	91

LIST OF TABLES

Table		Page
1.	Means and Standard Deviations: Number of Correct Recognitions..	77
2.	Context by Distract by Prompt ANOVA Summary: Number of Correct Recognition Responses...............	79
3.	Means and Standard Deviations: Confidence Ratings.....	85
4.	Context by Distractor by Prompt ANOVA Summary: Confidence Ratings....................................	86
5.	Means and Standard Deviations: Weighted Confidence Scores..	88
6.	Context by Distractor by Prompt ANOVA Summary: Weighted Confidence Scores............................	90
7.	Context by Distractor[a] by Prompt ANOVA Summary: Number of Correct Recognition Responses...............	129
8.	Context by Distractor[a] by Prompt ANOVA Summary: Number of Correct Recognition Responses...............	130
9.	Context by Distractor by Prompt[a] ANOVA Summary: Number of Correct Recognition Responses...............	131
10.	Means for Each Subject Used in Analyses Reported in Text..	133

ACKNOWLEDGMENTS

I gratefully acknowledge the contribution made by Jim Antes for his informed assistance and guidance throughout all phases of this investigation. As the chairman of my thesis committee, Jim willingly spent innumerable hours advising, encouraging, and prodding me through the less pleasant times involved in any project such as this one. I wish also to thank Mark Grabe and Paul Wright for their active participation on the committee, both offering helpful comments and suggestions at many junctures along the way.

Credit must also be given to Sue Burrus for the time and care she devoted to the preparation of many of the physical materials employed in this study. Thanks also go to those 92 individuals who served as raters and subjects. Their honest effort and cooperation allowed the experimental phase of this research to proceed smoothly and without hitch.

I reserve my most profound appreciation for my loving family. As a colleague, companion, and wife, Michelle has been an integral part of this thesis from its inception. Her support and faith in me have been unending, as were the many personal and professional sacrifices she endured to allow me to complete this project. Finally, I wish to acknowledge our three children, Abraham, Simon, and Zachary for the many intangibles that they have always given me.

ABSTRACT

Several investigators (e.g., Biederman, et al. 1974; Loftus & Bell 1975) have proposed that one's perception and subsequent recognition of pictured scenes results from the acquisition of two types of information. One is specific in nature and results from the direct inspection of object detail. The second type is more general in character and is thought to result from the processing of contextual information. Friedman (in press) has pointed out that context must be considered to encompass both internal (memorial) and external (physical) constraints, and as such to have a dual informational base. The present study was designed to assess the separate and combined effects of these internal and external sources of contextual information as they influence the amount of object detail later available to the observer.

Each of 72 (21 male) undergrads was presented with a written phrase prior to a 150 msec. exposure of a pictured scene, followed by a four alternative forced-choice recognition test. Subjects responded by selecting one of the four objects in the test as having been viewed previously, and then rated their confidence in that selection. The type of internal or memorial-based contextual information, prompted by a written phrase describing the theme of each stimulus picture, was varied within subjects such that each subject viewed one-third of the pictures preceded by a compatible theme, one-third preceded by a neutral phrase, and one-third preceded by an incompatible theme. External context and recognition test distractors were varied between

subjects. External or physical context was either present or absent, and the distractor objects were objects from dissimilar scenes, objects differing from those viewed but from similar scenes, or objects possessing the same generic name as the target but differing in some physical attribute.

Distractors from different scenes consistently resulted in the best recognition performance and the most confidence, with no differences in accuracy between the other two distractor types. Subject confidence was a more sensitive measure of distractor effects, as ratings were significantly higher when distractors were varied from the target in object rather than attribute information. Recognition accuracy and confidence were also enhanced when the pre-stimulus prompt was compatible with the stimulus picture, but only when distractor objects were from dissimilar scenes. A trend was evidenced suggesting that recognition accuracy may have been enhanced by the presence of external context when test distractors were from dissimilar scenes, whereas the absence of context may have facilitated performance when distractors differed from the target object only in physical detail.

These findings were interpreted via a consideration of the influences of the two types of information available in real world scenes and of the demands imposed on the subject by the particular task employed to study their perception and recognizability.

CHAPTER I

INTRODUCTION

The psychology of visual perception has evolved out of the efforts of scores of investigators who, over the course of many centuries, have tried to understand the ability of human beings to visually experience the world. The problem faced by each of these investigators has been the veracity of this experience; that is, the ability of people to perceive the organization and differentiation inherent in the environment, often on the basis of little information. During this evolution, much of the theoretical work has been founded upon the study of pictures and pictorial representations. Despite assuming a variety of forms, these representations generally result in a veridical experience comparable to that resulting from direct visual contact with the world. An explanation of the nature of this experience, and the structures and processes which are responsible for its occurrence, continues to be the challenge to those psychologists interested in how people perceive pictures.

As the study of picture perception has grown, two major theoretical frameworks have developed to explain present findings and generate questions for the future. Both advance explanations for the ability of people to perceive the visual environment in general, and both give special attention to the perception of its surrogates, pictures and pictorial representations. (For present purposes, pictures and pictorial representations are both included to cover the entire spectrum of visual surrogates, with pictures referring to high-fidelity representations such as photographs, line drawings, paintings, etc., and pictorial representations referring to more abstract and symbolic representations as might be exemplified in geometric figures or the sign-laden graphics of ancient cultures.) The two theoretical positions differ in orientation and focus of study, and with respect to the types of processes considered critical to an understanding of picture perception.

The first framework employs a "registration" approach in its study of visual perception. Its advocates, J. Gibson (1950, 1966, 1971), E. Gibson (1969), Kennedy (1974), and in his recent work, Turvey (1975, 1977), emphasize the importance of the information inherent in the visual environment. J. Gibson (1950, 1966) proposed that objects in the environment structure the light they reflect such that information concerning the identity, shape, size, color, and position of the object, and its relationship to other objects, is conveyed directly to the observer. With each glance at the environment, part of this available information is registered by the observer and, over successive glances, the information from each overlaps to impart the impression of a stable, solid, and complete visual world. The perception of this visual world is accomplished by attending to the invariant information existent within the variant information that results from successive glances at the same environment, thus precluding the necessity of assuming any mental structures to integrate the information from successive glances (J. Gibson 1966). The observer is assigned the passive role of registering the information crucial to an accurate perception of the environment (i.e., the invariant information), and assumes an active role only in searching the environment for information and continuously developing more refined abilities to discriminately attend to the most informative environmental stimuli (E. Gibson 1969). In the sense that the observer simply registers invariant information, this framework proposes that visual perception is a direct, ordinal process (Turvey 1977).

The registration approach contends that pictures convey the same information as the objects and scenes they represent, and are therefore perceived in an identical manner . . . the invariant information contained in the representation is registered over the course of several fixations upon the picture (J. Gibson 1971; Kennedy 1974). In addition, information is registered telling the viewer that his or her perception is of a surrogate, thus allowing its distinction from the direct perception of a real world object or scene.

The second major theoretical framework advances a "construction" approach to the study of visual perception. Its proponents, Hochberg (1968, 1970, 1972, 1978), Neisser (1967, 1976), Bruner (1957a, 1957b, 1960), Gregory (1970), and Vernon (1952, 1955, 1970), emphasize the cognitive activities of the perceiver and the interaction of these activities with the visual input from the environment. Based upon the notion that visual perception results from the products of looking behavior, Hochberg (1972, 1975) proposed that such perception involves purposeful, goal-directed behaviors. To accomodate the executive functions necessitated by such behavior, a mental structure is assumed that incorporates various cognitive components and functions actively to categorize (Bruner 1957b; Vernon 1970), select (Bruner 1960; Hochberg 1968, 1978; Neisser 1976), and test (Bruner 1957a; Gregory 1970; Hochberg 1970) input, and generally to serve the perceiver in the process of synthesizing visual information into a viable perception (Neisser 1967). Thus, in this framework, the observer is assigned an active role identifying and comparing visual input within the context of a memorial network, as well as gathering additional input to enhance the comparative process . . . in short, the role of constructing the perception responsible for one's visual experience of the world (Vernon 1952, 1955).

The perception of pictures and pictorial representations results, according to Hochberg (1972), from the integration of local features obtained from momentary glances and information already contained in the schematic map of the observer. The schematic map is the mental structure that is actively constructed out of information retained (and organized) from past experience and information presently being obtained from looking behavior. It is this schema or schematic map which functions in an executive capacity (though it is not necessarily synonymous with <u>the</u> executive) to guide the location of future glances, facilitate the encoding of information from the current glance, and as the standard or expectancy against which incoming information from local features may be tested for conformity (Hochberg 1968, 1972).

A primary difference between the two frameworks outlined is the postulation by the construction model of a dynamic mental structure or representation to function as the mediated perception of input from the environment; this mediated perception is the perception experienced by the observer at a given moment in time. The present study follows from this construction approach and assumes the existence of a mental representation active in the perception of pictures.

As background, the history of this notion of an active mental representation and a representative sample of experimentation on picture perception which evidences selective attentional patterns consistent with a model of mediated perception will be reviewed. This is followed by the presentation of a contemporary theory of picture perception, and subsequent pertinent research on the processes involved. The chapter closes by noting the implications of this theory and research as they relate to the current investigation.

A History of Process Mediation

The idea that visual perception involves some kind of mental representation has its roots in the early attempts of scholars to understand how our perceptions of the world are structured into organized and meaningful experiences. Although Locke had emphasized the importance of experience and Kant had proposed innate categories of understanding or pure intuitions, it was not until Helmholtz that a theorist effectively and cogently integrated their notions and expounded on the influence of an organized recollection of past experience on present perception.

In implicating the inferential nature of man, Helmholtz (see Boring 1950) suggested that current perceptions result from the organization of sensory input into familiar categories. This organization was seen to reflect the ways in which the regularities of the environment structure incoming information and allowed the utilization of capabilities for judgment and inference in the perceptual process.

Current perception was thus viewed as a combination of representations based upon past experience (i.e., familiar categories) and present sensation; this combination being the result of a complex, integrative, and inferential process. A perceptual experience was seen as the conclusion of this process, and synonymous with it (Pastore 1975).

James (1890) and Titchener (1919) thought perception was the product of accrued ideas and images as they interacted with present sensation, with the latter theorist emphasizing the importance of an integrated context as the basis for the meaningful (as synonymous with structure) quality associated with perception. Titchener related context to some nebulous mental constellation resulting from this accrual of ideas and images, and thus set the stage for the introduction of the construct "schema" to capture the envisioned content and process thought necessary for perception.

The term schema was first employed in the explanation of perception by Head in 1920, and has since revolutionized scientific theorizing on perception. (Actually, Kant used schema to denote his concept of pure intuitions as they functioned in perception, but clearly his usage has little relationship to the meaning given the term by theorists to follow [see Oldfield & Zangwill 1942-3]). Head introduced the notion of schema to account for the loss of postural perception suffered following lesions to various parts of the motor cortex. According to Head, this schema existed in the brain as an organized, constantly evolving model or standard against which current postural position could be evaluated . . . the resulting evaluation equivalent to a conscious perception. Thus, perception was viewed as the conclusion of a comparative process, a comparison between past impressions and current sensations made possible by the existence of this dynamic organization called a schema.

Bartlett (1932) made heavy use of Head's notion as an explanatory concept in his classic exposition on memorial processes, Remembering. He substantially enlarged upon the applications and functions of schemata, formally proposing them as influential and ever-present mediators in thinking, remembering, and perceiving. According to Bartlett, "'Schema' refers to an active organization of past reactions, or past experiences, which must always be supposed to be operating in any well adapted organic response" (p. 201, 1932).

His presentation of a mediating representation differed from Head's in at least three respects. First, Bartlett emphasized the behavioral aspects of schemata, proposing the organization of past reactions rather than past impressions. Second, Head restricted the occurrence of schemata to higher (cortical) levels in the nervous system, whereas Bartlett conceptualized their formulation at all levels. And third, Head frequently used the term "model" in his account in a manner suggesting he regarded schemata as situation-specific, whereas Bartlett placed great emphasis upon the generalized nature of schemata. This generalized character was seen to make our current experiences prototypical in nature, as opposed to specific (Bartlett 1932). Bartlett also went beyond Head in arguing that our constructed perceptions incorporate personal interests, values, and needs, and by introducing the notion of "orientation." Orientation was regarded as one of the primary functions of schemata . . . a function that made certain perceptual responses more probable in a given situation, such that incoming information is restricted in its impact upon the final perceptual experience.

Following Bartlett, numerous investigators (e.g., Ittelson 1954; Northway 1940; Wolters 1933) espoused active cognitive mediation of visual perception, most employing a schema somewhere within their particular theoretical framework. Generally, these theorists had little to add to the construct as conceived by Head and amended by Bartlett. There is, however, the work of a few investigators that merits discussion for its enrichment or new look at the function of a process mediator in perception.

An aspect of schema poorly developed by Bartlett is its function as a plan for perceptual behavior. It was Piaget in his voluminous writings on developmental epistomology who first emphasized this quality. According to Piaget (1954, 1973), schemata are by nature sensory-motor plans, cognitive structures related to classes of action-systems relating recurrent situations to a disposition to act. Schemata were seen to structure incoming information via the processes of assimilation, and were modified by accomodating to this information. Thus, perception was viewed as the product of this ongoing interplay between assimilation and accomodation. The planning character evident in Piaget's approach embraces the same notion of goal-directed behavior that is later found in the works of Miller, Galanter, and Pribram (1960) and Hochberg (1972, 1975), all of whom emphasized the importance of such behavior for perception.

The emphasis upon personal history, values, interests, and needs as they influence perception was enhanced by the Transactionalists and their descendents in the New Look (see Avant & Helson 1973). The

Transactionalists viewed perception as the product of many transactions between the individual and the world . . . transactions giving rise to assumptions regarding the nature of the world. These assumptions, heavily influenced by individual values and interests, are organized into a schematic representation of the world that functions to integrate current sensations into a meaningful perception. While proposing cognitive mediation in the form of an "assumptive world," they failed to address its nature or the processes responsible for its occurrence (Pastore 1975). It was the work of investigators such as Bruner and Vernon that would throw more light on the influence of these idiosyncratic factors in the categorization and integration of sensorial information.

Bruner (1951, 1957a) proposed that visual perception results from the matching of sensory information about an object to a cognitive "category." Stimulus information from the environment functions as a cue (see Brunswik 1956) used by the observer to infer category membership. Bruner's categories are analogous to schemata, functioning to organize object information within the currently existing network of information already present in the observer. These categories serve to guide the selection of cues in accordance with currently operating hypotheses regarding the nature of the object under consideration. It is the hypothesis of the observer that determines the accessibility of categories and incorporates individuals' needs, values, and personality patterns. Thus, "perception involves an act of categorization" (p. 123, Bruner 1957a), followed by the formulation of hypotheses used to select more information, and concluding with the confirmation of these cues as appropriate to the categorization. Idiosyncratic factors function to tune the hypotheses in favor of the individual (i.e., make them compatible with both past experience and present needs and interests), with the result that perception tends toward the typical or expected and may even distort the true classification of object information and result in a non-veridical perception (Bruner 1951).

Vernon (1955) equated assimilation and schematization as processes responsible for the integration and organization of recorded sense data into a mental representation or set of expectations. Schemata were seen to function in perception to produce a condition of expectation such that the observer knows what to look for in the incoming flood of sensory data and how to handle these data . . . "how to classify, understand, and name them, and draw from them inferences that give meaning to percepts" (p. 186, Vernon 1955). The construction of schemata, like the formulation of Bruner's hypotheses, reflect the different experiences of different individuals, and the interests that led them to seek these experiences. She viewed the influence of idiosyncratic factors as both indirect and enduring in their effect upon selecting and classifying information.

Utilizing the hypothesis-testing approach of Bruner, Neisser (1967) proposed that perception is the end result of a preliminary analysis of the visual field followed by an active synthesis of information from those objects attended plus information retained from previous acts of attention. This constructive act of synthesis thus relies heavily upon both memorial and attentive processes. The entire perceptual process is thought to be cyclical in nature, whereby a schema of the present environment functions to direct exploration for information, resulting in the sampling of specific environmental information which, in turn, modifies the schema of the present environment . . . and so on (Neisser 1976). According to Neisser, the schema represents only one part of the observer's "cognitive map" (Tolman 1948) or "visual world" (J. Gibson 1950) and is thus narrower in scope than either. To the extent that it directs exploration it is termed an anticipatory schema (see Woodworth 1938), and to the extent that it exists in a modified form, relative to its form in previous phases of the cycle, it may be called an orienting schema. However conceived, the schema is seen basically as a plan of action for selecting, analyzing, and interpreting information received from the environment such that a meaningful perceptual experience results.

A quite current approach to process mediation in visual perception is based on the concept of "frames" introduced by Minsky (1975). A frame is similar to a schema in that it may be viewed as a generalized representation of the information acquired from past experiences relating to a given situation. A frame exists as a hierarchically organized network of nodes and relations, the top levels of which are fixed and represent general information always true about a particular situation. Information at these upper levels of the network may be thought of as definitive with respect to the particular scene or object that the frame represents. The lower levels of the network function as terminals or "pigeon-holes" (Broadbent 1971) ready to accept specific detailed information gathered from the observer's current interaction with a particular situation. Prior to or in the absence of detailed input, default information (i.e., prototypical knowledge based upon the aggregate of past experiences the observer has had with that situation [see Evans 1967]) occupies these terminals and can be used to fill or read in missing data (Minsky 1975). Related frames may themselves be organized into a network or frame system, also thought to be hierarchically organized (see Palmer 1977 for a discussion of hierarchical organization in visual perception).

Frames may be seen to differ from schemata in that they tend to be essentially static in character, existent as an outline waiting to be completed by sensory information. Perception consists of the activation of a frame and/or its corresponding frame system by the presence of sensory data or by the preparatory activities of the perceiver in anticipation of sense data (Friedman, in press).

Applying the concept of an active mental representation to the study of visual perception, Hochberg's theorizing has incorporated most, if not all, of the theoretical work on schema since its conception, making his theory useful to consider as a summary to this historical presentation.

Hochberg (1975) asserts that the postulation of a schema or schematic map as a theoretical construct is necessary to account for the integration of information obtained from successive glances at the visual field because such glances may be separated by considerable time and space, and because some kind of selection occurs during the interval between glances. He proposed that schematic maps have the following functions: (a) expectancy generalization, (b) information generalization, (c) feature storage, and (d) peripheral selection of new information (Hochberg 1968). Schema thus serve to guide eye fixations via expectancies and peripheral selection,encode information obtained during the course of a fixation using the prototypical information held in a schematicized deep structure, and store the encoded information against which to compare new information for recognition and as a basis for a judgment about where to locate the next fixation. The generalized or prototypical nature of information stored in the schematic map is viewed by Hochberg as the reason the perceiver can go beyond the information given and generate information consistent with past experience and enduring dispositions. In emphasizing the purposeful quality of the looking behavior that ultimately results in a visual perception, Hochberg (1970) refers to schematic maps as "perceptuo-motor analogs" generated by the perceiver and used to extract and edit the most useful information from that available so that the experience is an organized and meaningful one. Coming to a meaningful experience is accomplished through the integration of successive samples of environmental information and is most efficient when the selection of information to be sampled is consistent with the currently operating schematic map. Therefore, schemata become a primary determinant of the allocation of attentional effort.

In concluding this review of the historical development of the concept of mental structures or schemata as active mediators in visual perception, it is important to note that there has been no attempt to discuss those theories (e.g., Hebb 1949; Koffka 1935; and in some respects Neisser 1967) which suppose mental representations or images possessing a one-to-one correspondence with the object or scene they represent (see Anderson 1978; Pylyshyn 1973; Sloman 1971 for such a discussion). The structure generally and variously referred to as a schema, schematic map, or frame is presumed here to have no such correspondence (i.e., it is propositional rather than analogical in character--see Palmer 1975b; Winston 1975).

For the most part, present usage of the concept of a mental representation entails a synthesis of Hochberg's and Minsky's descriptions, combining the emphasis of the former on process and the focus of the latter upon structure. This representation is assumed to be the construction of the observer built from various sources of information and functioning as a continually active process in the selection of information and in the modification of its own form. It is this representation that is perceived and is synonymous with one's perception at any given moment.

For clarity's sake, the term schema will be used in the remainder of this report whenever possible. However, the reader should remain alert to the subtle differences in meaning that have been ascribed to different terms used to construe mental representations involved in visual perception, and know that such differences may influence one's thinking appreciably.

Looking at Pictures

It is important to examine, at least in a representative fashion, prior studies conducted on attentional patterns in the perception of pictures and pictorial representations for at least two reasons. First, these studies provide evidence that the object of attention at any given moment is not the result of some random process, but rather reflects systematic and directed behavior, and thereby implicate the presence of some organizing and guiding structure at work during the perceptual process. And secondly, it is a reasonable assumption that the information extracted from a pictorial stimulus is highly correlated with the focus of attention (i.e., we extract more information about those objects and scenes to which we attend than about those to which we don't attend). Therefore, an examination of attentional patterns of subjects confronted with pictorial stimuli not only supports the notion of an active mental representation, but also yields great insight into the kind of information encoded by the observer and available for later use.

To address the question of attention allocation when viewing pictures, one may begin with the first systematic study of how people look at pictures. In 1935, Buswell used corneal reflection techniques and movie camera recordings to determine the eye movements of adults as they viewed numerous pieces of classical art. He discovered that two basic patterns of looking occurred: the first was described as a general survey of the picture marked by a series of short pauses over the major portions of the picture, followed by a detailed study of the picture involving long fixations concentrated over small areas of the picture. Buswell also noted that the duration of eye fixations increased as subjects continued to view the picture, suggesting that they were spending more time processing the detail of the picture as viewing progressed.

More recently, Yarbus (1967) found that people tend to direct their attention to the most informative aspects of a picture. Under free-viewing conditions subjects would fixate the eyes, nose and mouth of a face and would direct their gaze toward people rather than inanimate objects. Subjects also tended to fixate contours and other physical aspects of the picture likely to inform them of its content. Further, Yarbus found that subjects' pattern of eye movements exhibited a noticeable degree of repetitiveness over the course of viewing, with later fixations re-examining portions of the picture fixated earlier (see also Noton & Stark 1971).

Under more precise conditions, Mackworth and Morandi (1967) employed a free-viewing paradigm where the subjects scanned a picture for 10 sec. The major difference between their procedure and that used by Yarbus was that pictures were divided into 64 sections and each section was subjectively rated for recognizability and informativeness without the raters viewing the intact picture. Mackworth and Morandi found that the location and density of eye fixations of a different group of subjects viewing the entire picture were highly related to these ratings, with high informative sections receiving the most fixations and the longest duration. Analysis of each 2 sec interval showed that the density of fixations did not change with time, a pattern inconsistent with that found by Buswell (1935). Because their subjects immediately fixated informative (unpredictable) aspects of the picture, Mackworth and Morandi suggested that subjects might use information gathered peripherally to locate future glances.

A study by Antes (1974) further explored this tendency of individuals to focus upon the informative parts of a picture. Using stimuli divided into meaningful sections and subjectively rated for their informativeness, an independent group of subjects viewed 10 pictures taken from the Thematic Apperception Test for 20 sec under free-viewing conditions. He found that the density and duration of fixations, as well as the extent of saccades, were not consistent throughout the viewing period. Over the course of viewing the picture, subjects initially fixated for short periods of time on highly informative areas of the picture and gradually shifted to a looking response where fewer saccades (longer durations) were made and fixations were directed at the lesser informative picture parts. Consistent with the findings of Buswell (1935) two distinct patterns of viewing were evidenced. The initial and immediate attention to informative picture sections in this study provided further support for the notion of some extrafoveal mechanism guiding fixation location, and subsequently the extraction of information.

The controlled attentional response evidenced in the studies above is also found under task-imposed conditions. Yarbus (1967) had subjects view pictures under six different sets of instructions (e.g., evaluate the economic status of the family pictured), and noted that the density and duration of fixations was greatest on those aspects of the picture logically most informative with regard to the question posed by the experimenter (e.g., furniture, clothing, and other material possessions).

During a visual search task an individual's fixation and duration patterns are influenced by a number of variables. Gould (1967; Gould & Dill 1969; Gould & Peeples 1970) conducted a series of experiments where subjects were shown a nonsense pattern (the standard) in the middle of a display and were required to find its match among numerous patterns located in the periphery. Results revealed that increasing the similarity of target and non-target patterns lead to more and longer fixations, and that increasing the number of target elements had the same effect. It might be concluded that the degree of difficulty in a discrimination task is related to the probability of attention being directed to areas most helpful for the task's completion. Relevant also is the evidence that peripheral information is active in the determination of fixation location.

In another study of search times, Pollack and Spence (1968) investigated the impact of informativeness upon locating a targeted section from a picture. Five pictures were divided into 70 sections and rated on a 12-point scale for informativeness in terms of the importance of the section to the overall meaning of the picture. Assuming attention is necessary for rapid search, they found the pictures rated most informative to be discovered more rapidly than those rated least informative in all three search conditions they employed. Along with making fewer errors, these faster search times for highly informative traget sections provide additional support that attention is attracted by the most informative areas of the picture first.

The rated informativeness of picture sections as an intervening variable in recognition and localization tasks has also been studied. Antes (1977) divided color pictures into eight sections and had subjects rate the sections as either high or low in informational value with respect to the overall meaning of the scene. Using an independent group of subjects, he found that both recognition and localization accuracy depended upon the rated informativeness of the section probed and its location relative to the center of the picture. High informative and centrally located sections were more accurately recognized and located than low informative and peripherally located sections, with the effects of informativeness and location apparently independent.

Metzger (unpublished doctoral dissertation, 1976) also employed color pictures divided into eight sections, but sections were rated as high, medium, or low informative. Like Antes (1977), Metzger found that recognition depended upon the rated informativeness of the target section and its location in the picture. In addition, he found that informativeness and location interacted, with medium informative picture sections better recognized at peripheral locations while highly informative sections were more accurately recognized when they were centrally located.

Two recent studies have addressed the nature of informativeness in pictures. Taken together they provide support for the idea that the constitution of informativeness, and therefore a primary determinant of attentional allocation, involves both physical and cognitive components.

Antes and Stone (1975) employed a multidimensional similarity analysis of 10 judges' ratings of informativeness for a single stimulus picture divided into 32 sections. That analysis revealed that ratings of high-informativeness were primarily based upon the presence of identificable features or detail in the rated section, and the meaningfulness of these features or objects in terms of the meaning of the intact stimulus picture. That readily identifiable objects define highly informative picture sections has been confirmed in a study by Antes, Singsaas, and Metzger (1978).

Examining the determinants of eye fixations during picture viewing, Loftus and Mackworth (1978) differentiated between what they termed "physical" and "semantic" factors in characterizing informativeness. Cognitively, an object is informative to the extent that the object has a "low a priori probability of being in the picture given the rest of the picture and the subject's past history" (p. 104). Physically, an object is informative to the extent that it is non-redundant in an information-theoretic sense (see Berlyne 1960; Garner 1962). Subjects were shown 78 pictures for 4 sec with one-half of the pictures containing a cognitively informative object (e.g., a tractor in an underwater scene) and the other half containing an uninformative object (e.g., the same tractor in a farm scene). Results showed that subjects fixated earlier, more often, and for longer durations on the informative objects. In addition, the extent of eye movements to informative objects was relatively long (7 degrees), again suggestive of a peripheral editing process functioning to guide fixations to informative parts of the picture. Such a process is similar to that proposed by Mackworth and Morandi (1967) but is seen as dependent upon cognitive rather than physical features.

As noted at the outset, this presentation was intended to be representative rather than exhaustive. Numerous additional studies (e.g., Berlyne 1960; Day 1964; Mackworth & Bruner 1970; Williams 1966) could be cited that have investigated the variables that influence the allocation of attentional effort when viewing pictorial stimuli. Excellent reviews have been compiled by Rayner (1978) and Kahneman (1973). It might also be noted that a review by Wachtel (1967) details the influence of the observer's "style of approach" to pictured information, a notion which reflects the intensive (Berlyne 1960; Hochberg 1972) rather than selective character of looking behavior.

The studies reviewed in this section convincingly demonstrate that the sampling of pictured information is indeed controlled and not random. Under both task-imposed and free-viewing conditions, people look initially to the most informative areas of a picture. Most probably, such direction and control is the product of an active mental representation of the picture enhanced by information gathered from both foveal and peripheral vision. As viewing continues, lesser informative picture parts of the picture are examined and processed.

Regarding the information extracted as a result of such looking behavior, it may be inferred from the two distinct patterns evidenced in the work of Buswell (1935) and Antes (1974) that more than one type of information is being gathered. In addition to the detailed information gathered during each fixation on the picture, the initial broad sweeps across the picture with short fixations would present the observer with a diverse sample of the picture's content, and give the observer more general or global information about the picture. This idea of two kinds of information extracted during the perception of pictures has recently been developed into a general theory of picture perception, and is the topic of the next section.

A Duplex Theory of Picture Perception

As noted by Hochberg (1968, 1970), the postulation of some organizational structure is necessitated by the purposeful, goal-directed nature of our attentional response. In addition, the influences of past experience and momentary task demands require that such a structure be amenable to change while maintaining its identity. Classically, these requirements have been met by the construct "schema." As evidenced in the previous section, attentional patterns in picture perception also reveal this controlled and task-directed selectivity, and would thus also imply the existence and functioning of a schema.

A general theory of picture perception has recently been advanced that incorporates this idea of a schema. The theory is founded primarily in the empirical work of investigators interested in the effects of context on the perception of pictures, the most notable of which is Biederman. It was Biederman (1972) who first proposed that two types of information are active in the perception of pictures, but the antecedents to this notion are easily recognized. As reviewed earlier, Buswell (1935) identified what he thought to be two attentional patterns in his subjects' responses to classical art, and suggested that the long fixations over small areas of the pictures were indicative of the viewer processing detailed information about its content. The results of several studies (e.g., Hochberg 1968; Mackworth & Morandi 1967) support the operation of some kind of peripheral editing process prior to the focusing of foveal attention, and presumably based upon information different in character from that gathered foveally. Further, Karpov, Luria, and Yarbus (1968) remarked on the presence of what seemed to be two distinct processes of information encoding from their research with brain lesion patients. And, the distinction by Neisser (1967) between pre-attentional perceptual processes and processes involving perceptual synthesis suggest the processing of two types of information in the perception of pictures. This research notwithstanding, Biederman's research is the proper starting point for what might be termed a "duplex" theory of picture perception.

Because of the empirical genesis of this theory, Biederman's research will be presented first, followed by a summary describing the characteristics of the theory. Refinements and supporting evidence will conclude this section.

In his 1972 study, Biederman's subjects viewed 96 real-world scenes and were asked to identify which of several objects was in a cued position within the scene. He varied exposure duration (300, 500, and 700 msec), cueing order (before and after stimulus presentation), and whether the scenes were presented in a coherent or jumbled state. Jumbling was employed to remove the effects of context by destroying the information derivable from given spatial relationships among objects, and was accomplished by cutting each picture into six parts, rearranging them, and photographing them for presentation. In addition, Biederman presented the four response alternatives either before or after the presentation of the stimulus. As expected, recognition accuracy was best at all durations when the cue and response alternatives were presented beforehand and the scenes had been viewed in a coherent state. More importantly, he found a significant reduction in recognition accuracy when scenes were jumbled, even when subjects knew what to look for (alternatives before) and where (cue before).

These jumbling effects led Biederman to propose the existence of two "functional units" (corresponding to information types) involved in the perception of a scene . . . one based on individual objects, and the other a more global schematic type. In addition, he raised the question of the location of context effects, which he supposed corresponded to schematic information, in the processing sequence.

To examine further the impact of context on the perception of pictures, Biederman, Glass, and Stacy (1973) studied the effects of object probability and coherency on a speeded search task. Again, real-world scenes were employed so that the meaningfulness normally encountered would be maintained in the experimental setting. Before viewing the scene for 5 sec in either jumbled or coherent form, subjects viewed an object which they were told might be in the scene they were about to see. When presented with the scene they were to press either a "yes" or "no" finger key to indicate whether the object they had seen beforehand was contained in the scene. On one-third of the trials the object was from the scene. On another third of the trials the object was not in the scene but possibly could have been (e.g., an automobile followed by a street scene). On the last third, objects were not contained in the scene and their inclusion was very improbable given the meaning of the scene (e.g., an automobile presented before a kitchen scene).

Consistent with the 1972 study, jumbled stimuli required longer response times than coherent scenes, with the difference being most pronounced when the response category was "possible no." Biederman, et al. interpreted their results in terms of a schema model of processing, whereby an initial holistic or semantic interpretation is constructed followed by a more detailed analysis. Reaction times were slowest under the jumbled "possible no" condition because the initial holistic characterization, once achieved, did not allow the immediate recognition that the target object probably was not contained in the scene presented. The effect of jumbling in the "impossible no" condition was to require additional time for the construction of this overall, semantic interpretation before detailed analysis and comparison of the target object with those held in memory could be accomplished. Formation of this holistic characterization, though delayed, would be sufficient for rejection of the targeted object in this condition, and thus response times were faster for "impossible no" when compared to "possible no."

Important to the discussion of later studies, Biederman, et al. (1973) also noted that these effects may have resulted from some interference with the object identification process. And, by proposing that a holistic characterization is dependent upon a physical context, and that contextual information is the first to be extracted from a pictured scene, these authors suggest that the effect of context in the processing sequence is primary, and thus answer the question posed in 1972.

In a third study, Biederman, Rabinowitz, Glass, and Stacy (1974) limited the exposure duration of the stimulus picture to a maximum of 300 msec so that the information extracted must result from a single eye fixation, and studied the ability of subjects to select from among two verbal labels the one that accurately represented the theme of a pictorially presented scene. The stimuli used in the previous studies were again used, as was jumbling because of the belief that the theme of a scene is dependent upon the spatial relations among objects within it.

In the first experiment reported, they studied the effects of jumbling upon scene characterization. Jumbling was hypothesized to reduce subjects' conceptual accuracy (labeling) when labels were similar (i.e., where objects would be of less informational value). Scenes were presented for 20, 50, 100, and 300 msec. Jumbling, label similarity, and duration were all significant, as was the duration by label similarity interaction. The correct label was selected most often when labels were dissimilar and when scenes were presented coherently. In addition, these effects reached an asymptote by 100 msec, producing a ceiling effect probably responsible for the single significant interaction.

In the second experiment reported, the 1972 study was replicated except that durations were now 20, 50, 100, an 300 msec, and half of the subjects made selections from verbal response alternatives rather than pictorial ones. Scene version and duration were significant as was their interaction, with coherent scenes resulting in a higher percentage of correct object identifications at all durations, but with the difference in performance on coherent versus jumbled forms less pronounced at the longer durations. This improvement of the jumbled condition with increased exposure evidenced the impact of specific object information in overcoming the interference effects of jumbling as viewing progressed.

Again citing a schema model of processing, Biederman, et al. summarized their 1974 findings as reflecting the extraction of two kinds of information simultaneously. As specific object information is sampled from a scene, a holistic characterization including a thematic component is constructed. Both specific information and the more global information of the holistic schematic representation are available after brief 100 msec exposures, and each may be evidenced under the appropriate task conditions.

Taken together, Biederman's three studies demonstrate that the effect of a coherent context (i.e., the natural spatial relationships among objects, and the relationship between objects and the scene in which they are embedded) on recognizing objects from a scene is a facilitative one. Facilitation is greatest when the scene is viewed for a very brief time. And further, a coherent context helps the observer to semantically label the scene.

It is appropriate at this point to clearly describe the theory of picture perception which has emerged from Biederman's work. Basically, it is proposed that two distinct types of information are available to the observer when viewing a picture. The first type may be characterized as detail or specific object information, probably acquired from foveal or near-foveal vision. The second type of information is more global or general in nature, and is probably founded upon information gathered both foveally and peripherally. That is, this global characterization of the scene is constructed out of information concerning the identity and spatial relationship of the scene's content.

The construction aspects of this theory are of paramount importance because the perceiver's realization (i.e., awareness) and use of global information is made possible by the existence of a schematic representation of the scene constructed out of contextual and specific object information, and tempered with the knowledge gained from past visual experiences with such scenes and held in memory. If this schematic representation is considered to be hierarchical and pyramid-like in structure (see below) then general information about a scene is most probably held in the upper or higher levels of this representation. It is further assumed that one's global interpretation of a scene contains a semantic component.

Both types of information about the pictured scene are acquired simultaneously. However, because there are innumerable pieces of detailed information potentially available in any given scene, the complete processing (i.e., extraction, encoding, storage, etc.) of this information type may begin simultaneously with the extraction of global information but will take considerably longer. Therefore, when a scene is viewed for a very brief period of time contextual information is the major informational source for the completion of most tasks. Further, when context is jumbled or absent, the extraction of global information in the form of a holistic characterization or scheme is interfered with, and the construction must then be based solely upon detail or specific object information as a result. With prolonged viewing, perception is based more upon the increasing amount of detail information being acquired. It might also be inferred that this detail is incorporated into the schematic representation of the scene so that after prolonged viewing one's mental representation, and thus one's perception, includes both a global understanding of the scene and an accurate list of its specific detail. This theory may be termed a "duplex" theory of picture perception because the information synonymous with a perception is of two types and is acquired simultaneously.

A theoretical work by Palmer (1975b) describes a possible structure for scene schema and in the process addresses how the global information is integrated and interacts with the detailed information gathered from a scene. According to Palmer, a schema may be described structurally as a set of relationships between various informational entities, with sets of information organized hierarchically. These entities contain sense data concerning the physical properties of objects at the lower levels, and exist as object schemata at higher levels. A scene is thus represented structurally by the relationship among various object schemata. Objects are represented structurally as a set of physical properties (e.g., size, color, orientation) occurring in a certain relation to one another. Applied to the duplex theory of picture perception, global information or a holistic characterization of a scene is analogous to that information contained at the upper levels of a scene schema. Detail information about an object is represented as one of a collection of values involved in the characterization of the object.

Noting that recognition is facilitated by a coherent context and when the stimulus is well known rather than novel, Palmer asserts that one's knowledge gained from past experience enters into the perceptual process in the form of a propositional schema as described above. The observer enters a scene with certain schemata primed to characterize that scene. The schema tentatively selected to facilitate the scene's interpretation is determined by the initial input from the gaze of the observer in the form of contextual information. Detail information is incorporated into the lower levels of whatever conceptual interpretation has been adopted. In this manner both types of information are processed simultaneously. When contextual information is not available, the selection of a conceptual interpretation is delayed and requires the trial and error processing of considerably more detail before one's final perceptual interpretation of the scene is possible. When a conceptual interpretation has been adopted, whether tentatively or otherwise, it causes the observer to seek or expect confirming sense data. Contrary data may be misinterpreted or cause a different conceptual interpretation (schema) to be adopted which is more consistent with the current input. Palmer's (1975b) structural and process analysis reflects the interaction between global and detail information which is of primary import in understanding the perception of scenes.

In addition to the research by Biederman (1972; Biederman, et al. 1973; Biederman, et al. 1974) presented above, support for the duplex theory of picture perception in the form of evidence for the existence and activity of a holistic characterization has come from a variety of studies.

Potter (1975) tested the notion that the meaning of a pictured scene is extracted very early in the viewing process. She presented subjects with either a target picture or a name for the target picture and then gave them a recognition test. Names given the target scenes were brief descriptions of the major objects or events pictured (e.g., a boat). During the test phase, pictures (the target and 15 distractors) were exposed for 125, 167, 250, or 333 msec. Potter reasoned that to rapidly detect a target defined by its meaning rather than a specific visual pattern, the subject would have to semantically identify the target scene as it was presented during the test phase. Her results showed that subjects were able to recognize a target scene as rapidly and accurately when they knew only its name as when they had viewed that scene in advance. This finding held even at the shortest exposure duration employed (125 msec). As predicted by the duplex theory of information extraction, a pictured scene is rapidly processed to an abstract level of meaning.

In a 1977 report, Biederman explored the semantic and syntactic properties of schemata in their influence upon scene recognition. Noting that schemata are effective within the first 100 msec of viewing a scene, he developed a technique to violate specific relations between an object and the scene within which it is embedded. The five relations studied were support, interposition, probability, position, and size. As the first two embody physical restraints on object-scene relationships, they were considered syntactic components. Probability, position, and size reflect referential meaning, and were thus referred to as semantic components.

Employing line drawings of scenes on acetate, 247 scenes were composed with the necessary violations and xeroxed for presentation. Scenes were presented for 150 msec followed by a cued object, and recognition was tested. In looking at the error rates, targets violating a relation were less accurately identified (yielding miss rates of 45% compared to 25% for target pictures with no violations in relations), and there was a tendency for misses to increase when several violations were included. There were no significant differences in the error rates for violations in syntactic as compared to semantic relations. In the violation detection task reported, the cue preceded the scenes and subjects were required to press either a "violation" or "no violation" key. Again, there was no difference between violations detected based upon the type of violation, but as the number of violations increased from one to three, the speed of detection also increased. In a third task reported, subjects were to respond when they detected a specified violation. In this task, reaction times were slowed when an irrelevant violation was included, also suggestive of the similarity of violation types.

Citing the duplex model, Biederman interpreted his results as reflecting the operation of a scene schema immediately upon the presentation of the picture. As the schema is initially constructed from the information gathered from a single fixation and primarily based upon the contextual information available, violation of this context via the destruction of specific relations also destroys or inhibits the proper functioning of the schema. Object identification is made more difficult (in terms of both time and accuracy) and the recognition or detection of a given violation is confused. Additionally, the violation identification task demonstrates that knowledge of physical or syntactic relations does not precede or succeed knowledge of referential or semantic relations . . . rather, both types of relations are active in the construction of a holistic characterization of the scene.

The Loftus and Mackworth (1978) study cited in the previous section also supports the idea that contextual information and its resultant holistic characterization is available to the observer early in the processing of pictures. Their subjects fixated immediately upon what they termed "informative" objects. As noted, what defined an object as informative was the context of the picture and the subject's past history. Therefore, for subjects in their study to select informative over uninformative objects for the focus of attention required that those subjects immediately process contextual information and form a holistic characterization of the scene. Loftus and Mackworth suggested that this process allows the observer to assimilate these improbable objects as part of the currently viewed scene with the scene schema carried into the situation remaining otherwise unaltered. After the first fixation when the holistic characterization is first generated, a comparative process begins whereby schema generated expectancies are tested against incoming sense data. This process occupies the remainder of the observer's picture processing time.

Support for the functioning of a holistic characterization in picture perception may also be found in Antes (1977) investigation of recognition and localization accuracy in two experiments. In both experiments photographs were divided into eight sections and rated by an independent group of subjects for their informativeness (i.e., the amount of information the section contained relative to the entire picture).

In the first experiment, pictures were presented for 100 msec followed by one of the sections serving as a probe or target, and subjects were to decide if the probe section had been part of the stimulus picture and assign confidence ratings to their decisions. Probe sections rated as highly informative were better recognized than those rated low informative, and sections occupying a central position in the stimulus picture were also better recognized. Using a signal detection analysis of errors it was discovered that low informative probes yielded more false alarms and high informative sections yielded more misses. Assuming fewer identifiable objects in the low informative sections (Antes & Stone 1975), it was probably necessary for subjects probed with these sections to rely upon whatever schema might have been derived from the contextual information in a single glance, and thus subjects would be more likely to respond positively when there was no match between probe section and picture. However, when sections were high in informativeness, and thus contained identifiable objects, such confusions were less likely, and negative responses when there was a match dominated. This interpretation is entirely consistent with the duplex model, and further suggests that the subject's schema or holistic characterization of the scene will be automatic in its effect upon recognition responses, utilizing whatever information is available (i.e., it is not a voluntary process that is engaged only when sufficient information exists to avoid most error).

In the second study reported, subjects were shown the stimulus picture for either 100 or 500 msec followed by a probe of one of the sections. Subjects were to indicate on a 2 x 4 grid where in the picture the probe section had been located. Both high informative and centrally located sections were better localized. While correct responses were made only 31% (100 msec) and 38% (500 msec) of the time, incorrect responses were often made to adjacent grid areas suggesting that the general location of the section had indeed been determined. As subjects tend to fixate the highly informative sections of the picture first, this overall poor localization accuracy prompted Antes to speculate that the underlying processes responsible for recognition accuracy and localization may be different. It seems plausible that detail information greatly enhances recognition performance but only indirectly aids the localization of an object by initiating or activating a schema of the scene. The spatial information necessary to localize objects is tied to this schematization which functions primarily on past experience in early viewing and utilizes contextual information as viewing progresses. Until the schema has had time to evolve and gather this external spatial information, localizability remains at a low level compared to recognizability.

Loftus and Bell (1975) designed a study to investigate the extent to which picture recognition is based upon specific versus global information. Subjects were shown 60 pictures followed by 120 pictures (including the original 60) as part of a yes-no recognition task. The pictures were either photographs, embellished line drawings, or unembellished line drawings. Five exposure durations were employed: 60, 100, 250, 350, and 500 msec. The relevant findings were that both the amount of detail contained in the picture and the exposure duration were significant in their effects upon recognition. Accuracy improved with increased viewing time, and photographs were better recognized than embellished line drawings, and these better recognized than unembellished line drawings. These findings were bolstered by consistent confidence rankings and subjective reports. Subjects were more confident when more detail was available, and reported more often that their recognition judgments had been based upon specific detail rather than on feelings of familiarity.

Loftus and Bell proposed that the additional detail afforded the observer in the richer (defined in an informational sense) stimuli provided a better opportunity to encode a specific detail, with the result of such encoding being a "quantuum leap" in the available global information. This is consistent with the duplex model which proposes that recognition will be best under conditions where both specific and contextual information are available. While it might seem reasonable to expect that if recognition of briefly presented scenes is based predominantly on global information then accuracy should be equivalent among the three types of stimuli used in this experiment since all afforded global information, the obtained results appear even more reasonable when it is remembered that both specific and global information are available to the observer and incorporated into the schema representing a scene. The scene having the most evolved or embellished schema will be more accurately recognized than one containing fewer specifics, therefore the better recognition of photographs in the Loftus and Bell study. The important contribution here is the emphasis upon specific information in the processing of pictures . . . an emphasis which is not at the expense of the part played by more global information in that same processing.

The interactive nature of specific and contextual information in the recognition of pictures was also evidenced in the study of Metzger (unpublished doctoral dissertation, 1976) cited earlier. His finding that medium informative sections of pictures were better recognized than high informative sections could be predicted on the basis of the content of medium informative sections. These sections contained both specific object information and contextual information, and as such contributed more to the scene schema than high informative sections where only specific information was available.

Support for and refinements of the duplex theory of picture perception may also be found in the study of context, which evidences the impact of contextual information in the formation of a holistic characterization of scenes necessary for their interpretation and perception. As the current investigation is itself primarily concerned with the study of context, an examination of these studies will be presented separately in the next section.

In concluding this section, it might be noted that the present study adheres to the duplex theory as set forth above. When an observer is confronted with a scene, both specific and global information are available and are active in creating his or her perceptual experience. As indicated in this and the previous section, detailed information is acquired through the process of foveal sampling of the scene's highly informative sections (and medium informative sections to the extent that they contain readily identifiable objects). Global information is gathered primarily from the context of the scene via peripheral sampling of both medium and low informative areas, but is also based upon the results of comparative operations which have occurred previously and involve the relational characteristics of objects and scenes. These two types of information interact with each other over the course of viewing and perceiving the scene. Initial viewing activates a comparison between expectations held in a schematically structured memory and those acquired from contextual information with incoming detail information. The outcome of this comparison process entails that future scans of the scene be directed to areas that might confirm these expectations or resolve any conflicts between expectations and input. As in the cyclical processing advocated by Neisser (1976) and Hochberg (1970, 1978), the specific information gathered from these discrete searches will modify the scene's conceptualization, leading to altered expectations, altering the outcome of various comparisons, and ultimately changing the conceptualization of the scene synonymous with its perception.

Contexts Effects in Picture Perception

What are the effects of contextual information on the perception of a scene and the objects embedded within it? Despite the efforts of many investigators, this question remains largely unanswered. What is known concerning context effects in the perception of pictures has emerged primarily from the work of Biederman (1972, 1977; Biederman, Glass, & Stacy 1973; Biederman, Rabinowitz, Glass & Stacy 1974) and others to be presented in this section who have studied the perception of pictures directly (i.e., have used pictures as stimuli). However, relevant findings have also come from investigators employing non-pictorial stimuli (e.g., words, geometric configurations) and deserve mention here.

A pioneering study into the effects of context on the recognition of words was conducted by Tulving and Gold (1963). A total of three experiments were reported, all stemming from the basic premise that the amount of stimulus information required by a subject to recognize a word varies inversely with the amount of contextual information provided. In the first experiment, context length (8, 4, 2, 1, or 0 words preceding the target word) and the congruity of context (incongruous contexts were incompatible with the target word) were examined. An example of Tulving and Gold's stimuli is the following: "Three people were killed today in a terrible highway COLLISION" (p. 321), where the word in uppercase letters served as the target. Using the target word "RASPBERRY" within that sentence would constitute an example of an incongruous context. Ten target words were tested under the various combinations resulting from the two variables, and length of tachistoscopic exposure necessary for a veridical report of the target word was the dependent variable.

Results clearly showed that increasing the length of congruous context facilitates word recognition, whereas increasing the length of incongruous context interferes with recognition. Experiments 2 and 3 demonstrated a strong monotonic relationship between contextual information and length of context, and between congruity and length of context. In addition, they established that the dominant relationship was between context congruity and context length. It was therefore concluded that it is not the amount of contextual information but its relevance that accounts for the effects of context on word recognition. These results also clearly implicate contextual information in the activation of expectancies or hypotheses about the nature of the specific stimuli being observed.

In a study designed primarily to examine the issue of serial versus parallel processing (see Estes & Taylor 1966; Sperling 1967) Reicher (1969) measured recognition performance for one or two letters, four-letter words, and four-letter non-words. After tachistoscopic presentation of the stimulus followed by a 100 msec masking field, each subject was required to select the letter they had seen in the stimulus presentation from a pair of letters (all letters in the alphabet were sampled). In addition to stimulus type, Reicher varied exposure duration (60, 75, and 90 percent performance levels for a given individual) and cue presentation (response alternatives were presented either before or after stimulus presentation).

The major findings were that performance on one word was better than performance on either one letter or one quadrigram, and that performance on two words was superior to that on either two letters or two quadrigrams. This trend was unaltered by presenting the alternatives beforehand or by increasing the exposure duration. In addition, confidence rankings by the subjects were consistent with these findings; subjects were more confident on words than on the other types of material. This "word superiority effect" provides obvious support for the notion that information is processed in parallel. It also suggests that what is extracted from the stimulus presentation is the meaning of the stimulus, with the conceptualization faster for words than for letters alone or for quadrigrams in Reicher's study.

In 1972, Wheeler conducted a study that paralleled in essential detail of design the study of Reicher (1969). The major difference in paradigm consisted of the addition of several controls in the material used as stimuli (e.g., positional cerainty, testing of letters that also function as words in the English language). The persistence of the word superiority effect was evidenced, as Wheeler obtained results identical to Reicher's, and to a greater degree, 10% compared to 8% superiority of words over letters and quadrigrams.

Wheeler suggested that the recognition of words cannot be accomplished via independent letter recognition processes; rather, there must exist an interaction among the letters of a word such that a context results. It is not a contextual noise or narrowing effect, as poor performance on the quadrigrams indicated, but the context as meaningful information relevant to its constituent parts that is facilitative in the recognition task. Additionally, Wheeler proposed that his results are amenable to a simultaneous constraints model of processing, which he described as an initial parallel feature extraction stage followed by a second stage based on information gathered during the first and functioning to locate, construct and encode the stimulus. The consistency of this proposal with the duplex theory cited earlier is evident.

An investigation conducted by Weisstein and Harris (1974) examined the ability to recognize line segments embedded in various types of context. Four line segments differing in orientation and location within the visual field were presented at one of three exposure durations ranging from 5 to 44 msec. For a given trial, one of these lines was shown together with one of six contextual patterns. The subject's task was to view the stimulus (followed by a 100 msec mask) and identify which of the line segments had been presented.

The main finding was that when a line segment was part of a configuration (context) that appeared unitary and three-dimensional it was better recognized than when presented within a flat, non-meaningful pattern configuration. Weisstein and Harris suggested that this result with non-verbal stimuli is analogous to the word superiority effect found by Reicher (1969) and may be viewed as evidence for the efficacy of contextual information, especially meaningful contextual information, in the rapid perception of stimuli that are more spatial in nature than words or letters.

In a series of five experiments, Pomerantz, Sager, and Stoever (1977) investigated what they termed the "configural superiority effect." This effect is analogous to the word superiority effect, but arises when stimuli are geometric configurations rather than letters or words. In all the experiments reported, line segments of differing orientations, lengths, positions, and curvatures were used as stimuli.

In Experiment 1, subjects were presented with either a single parenthesis (no context condition) or a pair of parentheses (context condition) for 200 msec. When presented as a pair, only one parenthesis served as the target. The targets were varied to face either left or right (open or closed position), and the subject had to make a forced choice as to the direction of the target. Analysis of reaction times showed a significant difference, with times for the context condition being the fastest.

In the second experiment, Pomerantz, et al. minimized memorial involvement by presenting subjects with an array of 16 stimuli, four of which were different from the remaining 12, and required subjects to indicate the location of the disparate quadrant in the array. Additionally, this experiment controlled for the symmetry, spacing, and mirror-imageness of the stimuli. Results indicated that the configural superiority effect persisted with the additional controls. Also, it was found that arrays containing additional irrelevant contextual information led to poorer performance. Thus contextual information may facilitate or inhibit the speed of localization depending upon its relevance.

In the third experiment reported, the discriminations involved were of a broader ecological sample (e.g., orientation of curved line segments, position of a line relative to a fixed point, line length), and arrays were reduced in size from 16 to 4 elements so that individual elements or features might be enlarged (display size was kept constant). Analysis of reaction time measures showed that some of the discriminations were faster when made under context conditions, while others were slower.

In Experiment 4, subjects did not know in advance of presentation whether they would be viewing the stimulus in context or without. Stimuli were two arrays from Experiment 3, one with context and one without. Again, the configural superiority effect was evidenced and discriminations were faster under context conditions.

The last experiment Pomerantz, et al. reported was concerned with the serial versus parallel nature of processing from geometric configurations. Subjects were required to determine whether there were any disparate elements embedded in an array (i.e., a visual search task), under no context, good context (highly distinct and identifiable contextual elements), or poor context (vague and complex contextual elements) conditions. Array size (2, 4, or 6 elements), presence or absence of disparity, and location of disparity within the array were also varied. These last two variables were for control purposes only and were not analyzed. Reaction times were best for good context, followed by no context, and larger arrays were responded to more slowly. These findings were confounded by the interaction between the two variables and further analysis revealed that array size affected only the detection of stimuli embedded in poor contextual arrangements. Further, no configural superiority effect was obtained when the good contextual arrays were compared to those without context. However, Pomerantz, et al. proposed that this last finding may have indeed reflected a context effect in that the arrays without context, when taken as a whole, tended to form schematicized representations of letters of the alphabet, and were probably responded to as such. It was thus suggested that parallel processing had been evidenced.

The major conclusions to be drawn from the work of Pomerantz, et al. (1977) is that context improves discriminability under certain conditions and, on the basis of the oddity and search tasks which minimized the influence of memory, context has its effect on the perceptual rather than the post-perceptual processing of geometric configurations. The authors proposed an emergent-features explanation of the effects of context on perception, suggesting that the emergent features are more novel than the elementary features, and are thus more discriminable than features presented as targets without context. The inconsistent effect of context in these studies is probably due to the failure of a recognizable emergent feature to result from every contextual arrangement. Translated into the terms of the duplex theory, not every contextual arrangement results in the activation of a familiar conceptualization; the more foreign the contextual and specific information, the more active construction necessitated at the time of viewing for a coherent and meaningful experience to result. In proposing these emergent features in the perception of a superordinate spatial configuration, Pomerantz, et al. imply a direct perceptual access to this higher order information, a notion entirely consistent with the duplex theory presented earlier.

From these experiments employing non-pictorial stimuli, it is evident that contextual information plays an important role in the perception of a broad range of stimuli as measured by a variety of tasks. And while this review has not been an exhaustive one, the results of the studies presented here indicate that information derived from the context of a stimulus can either facilitate or inhibit performance on a task depending upon its relevance and meaning. Taken as a whole, these studies suggest that a higher order conceptualization of the stimulus may be derived primarily from the contextual information at hand, and that such context is an active informational source in the total processing sequence. Consistent evidence has come from studies employing pictures (e.g., photographs, line drawings) as stimuli.

In 1975 (a), Palmer investigated the effects of context on the recognition of objects from line drawings of scenes. Subjects were first shown a slide containing a contextual scene or a blank for 2 sec. Following a 1300 msec delay they were shown a second slide containing the target object for a duration of 20, 40, 60, or 120 msec. Subjects were to respond by writing the name of the object perceived, and to assign a confidence ranking to their judgment. For the first slide viewed, context was either absent, appropriate, or inappropriate. Within the inapporpriate context condition there were two subconditions: one where the following target object was similar in shape and size to an object appropriate to that scene, and one where it was dissimilar along these dimensions.

The principal finding was that accuracy of recognition was highest when the target was preceded by an appropriate context, less so when preceded by a blank slide, and poorest when it followed an inappropriate context. Furthermore, performance in the "inappropriate context--similar target" condition was inferior to that in the "inappropriate context--dissimilar target" condition. Increased exposure duration also enhanced performance. These findings evidence the importance of both contextual information and specific physical characteristics as they influence a subject's response. When subjects had contextual information with which to conceptualize object information they were better able to recognize or recall objects than if no contextual information was available. If inappropriate to that in which the object is normally encountered, that same contextual information may result in an incorrect characterization and thereby give rise to a non-veridical response. The incorrect characterization of an object is more likely when it physically resembles an object appropriate to another conceptualization, invoked here by an inappropriate context. That the inappropriate-dissimilar target condition resulted in better performance demonstrates that the conceptualization of an object is based upon physical or specific detail as well as upon context. The improved performance with increased exposure represents the impact of an increasing store of specific information about the object with longer viewing.

A study reported by Antes and Metzger (in press) looked at the effects of context upon the recognition and localization of objects in pictures. Two experiments were reported, the first being an extension of the works of Biederman (1972) and Palmer (1975a).

In Experiment 1, subjects were presented with 100 msec displays of either line drawings containing six objects embedded in the context of a scene, or drawings of those same objects in an array without the background context. This allowed for the reduction of context (rather than disrupting it as Biederman had done) and the presentation of contextual and object information within the same exposure (improving upon Palmer's design). Stimulus construction also allowed the objects in the array to maintain the same spatial relationships as they had in their contextual presentation. Following the stimulus presentation, subjects were required to select the object that they had seen in the display in a four alternative forced-choice recognition task. The alternatives were either consistent or inconsistent with the context of the scene in which the target object was embedded. Location of target objects within the display was controlled for by randomly assigning near and far objects from each stimulus to two trial series.

Antes and Metzger found that target objects were selected from inconsistent response alternatives most accurately and that target objects near the center of the display were recognized better than those located in the periphery. Analysis of the interaction between context and distractor consistency revealed that recognition was most accurate when objects were presented in context and the distractors were inconsistent with that context. Context also interacted with location, showing that objects presented without background context were recognized better if they were located in the central portions of the display. Antes and Metzger proposed that these findings demonstrate that context does not simply improve recognition for embedded objects; rather, it allows subjects to bias their responses in favor of objects appropriate to the scene. Context thus functions primarily to enhance the general characterization of the scene and facilitates object recognition only indirectly via inferential processes utilizing information contained in that characterization. Their results further indicate that contextual information is not the major informational source for the recognition of centrally presented objects.

In the second experiment reported, the effect of context on the ability to localize objects was investigated. The stimuli and their presentation was identical to that in Experiment 1. The task of the subject was to locate a cut-out of the target object within a frame equated in size to the stimulus display in the same position it occupied in the stimulus presentation. Analyzing the difference between the subject's positioning of the target object and its true location, and the difference between the distance of the target from the picture's center and the distance of the subject's positioning and center, they found that context was a significant factor . . . subjects in the context condition were better able to accurately localize the objects. The true location of the target object (central versus peripheral) also had an effect, with subjects placing near objects further from the center of the picture than they truly were, and inversely placing peripherally located objects closer to the center.

Antes and Metzger concluded that context facilitates the location of objects by providing information used to construct a representation of the scene which serves as a "frame of reference" for the spatial orientation of objects.

A study conducted by Antes, Penland and Metzger (manuscript in preparation) compared the effects of context on the recognition of both usual and unusual objects presented in pictures. For comparative purposes, the conditions of Antes and Metzger (in press) were replicated, except the target objects were now inappropriate to the context in which they were embedded. The subject's task was to identify the target object from among four response alternatives that were either consistent or inconsistent with the context of the picture. Because of the unusual character of the target objects, none of the objects available for selection in the inconsistent distractor condition were related to the context of the picture. To the extent that contextual information is utilized in the recognition of objects, this condition was expected to yield the poorest performance. Preliminary analysis showed that unusual objects were more poorly recognized when presented in context than when presented without background context. As expected, unusual objects were recognized at chance levels when subjects were tested under the inconsistent distractor condition. However, poorest performance resulted when the distractors were consistent with the context of the picture; a result consistent with the suggestion of Antes and Metzger (in press) that contextual information may bias responses in addition to altering the subject's perception and conceptualization.

Subjecting the combined results of this task and those of Antes and Metzger's first experiment to analysis revealed that object usualness interacted with both picture context and distractor consistency. The recognition of both usual and unusual objects was approximately equal in the no context condition, whereas the recognition of usual objects was significantly above chance in the context condition and the recognition of unusual objects was significantly below chance in the context condition. These findings clearly demonstrate that the ability to recognize an object is facilitated by contextual information if derived from an appropriate context and inhibited by information gathered from an inappropriate or atypical context. Antes, Penland, and Metzger (manuscript in preparation) suggest that contextual information is active during both the encoding and response periods in a study such as this one, affecting the encoding of specific object information such that only usual objects are encoded, and altering the subject's response such that it is directed toward objects consistent with the context of the picture.

Mandler and her colleagues have conducted a series of studies (Mandler & Parker 1976; Mandler & Johnson 1976; Mandler & Ritchey 1977) investigating the effects of context on the recognition of various types of pictorial information. Mandler and Parker (1976) presented four scenes (line drawings) which were either organized (objects were in their normal spatial relationship to each other and horizon markings were evident) or unorganized (essentially jumbled with horizon markings absent). Subjects viewed the pictures for 20 sec each with a 10 sec interval between pictures. They were then given an immediate recognition test followed by a localization task with the objects selected in the recognition task being given to the subject to locate within a frame. Mandler and Parker varied the size, orientation, and appearance of the pictured objects to yield an eight alternative forced-choice recognition test for each of the objects in each of the pictures.

None of the three variations employed resulted in differential performance on the recognition test for organized versus unorganized scenes. Apparently the context of the picture had little effect on such descriptive information. Analysis did reveal that picture organization had a significant effect upon localization performance; objects in organized pictures were localized more accurately than those viewed in unorganized pictures. Mandler and Parker's findings suggest that contextual information is more important to the spatial representation of objects than for their subsequent recognition. Keeping in mind the 20 sec exposure duration used in this study, such a finding would be expected given the amount of time available to gather specific object information. Both detail and more global information would be available in sufficient quantity after 20 sec and the former would be a far more accurate source for making the discrimination required in the recognition task, whereas the latter would be more useful in the localization task, probably because localizing an object within a scene requires some rudimentary characterization of the scene as a whole (in the relationships among objects in the scene, including their spatial relationship), and this is precisely the character of global information.

Also in 1976, Mandler and Johnson employed a same-different recognition task in the comparison of organized versus unorganized pictures, and their effects upon the type of information encoded. Their stimuli were 10 line drawings of scenes presented for either 5, 20, or 60 sec (held constant for a given subject). The alternatives in the recognition task were pictures identical to those used as stimuli, but having undergone one of five transformations: an object was moved, two objects in the picture exchanged places, an object was deleted, an object was replaced by another object of the same size and shape but conceptually different, or an object was replaced with an object of the same conceptual class but different in size and shape.

After converting their data to signal detection measures, Mandler and Johnson found that hit rates varied significantly for exposure duration only, with longer presentations resulting in higher hit rates. Analysis of correct rejections revealed more correct rejections for organized scene versions when two objects had been exchanged, and for unorganized versions when an object was moved or deleted. The other two transformations employed had no effect upon correct rejections when comparing organized versus unorganized presentations. As with hits, correct rejections increased at longer exposure times. Analysis of response latency revealed that all responses were more rapid to organized versus unorganized pictures with 5 sec exposure, whereas responses were slower for organized pictures for durations of 20 and 60 sec.

From the signal detection analysis comes evidence that context may be used primarily in the rejection of incorrect alternatives, rather than as a basis for a direct identifying response. This logically follows from the contention that contextual information tends to be global in character and, therefore, will not be useful to the observer in discriminating detail but will allow gross discrimination of the kind necessary to reject incorrect alternatives. As noted, hits did increase significantly with increased exposure duration, demonstrating the impact of an ever-increasing pool of detail or specific information. The effect of context in detecting exchanges between objects suggests that it plays a more important role in spatially representing objects in the scene, while the meaningful quality associated with the characterization presumably derived from contextual information inhibits the detection of changes in object density within the picture. The varying effects of scene organization in combination with exposure duration is more difficult to account for. At the 5 sec duration, the expected advantage of context is evidenced, while at longer durations latencies for the no context conditions were shorter. Mandler and Johnson suggest that detections were made equally fast at all exposure durations, but that during the longer exposures subjects spent more time exploring organized pictures subsequent to the detection than they would spend exploring pictures without context.

Mandler and Ritchey (1977) presented their subjects with eight slides of either organized or unorganized pictures for 10 sec each. After viewing all eight slides, they were to respond by making a same-different judgment and assign a confidence ranking to their decision. Recognition tests were given immediately, one day later, one week later, and after one month's time, however only the immediate recognition test is pertinent to this discussion. Eight different transformations and substitutions were made on the pictures used as distractors in the same-different recognition test. The relevant finding is that picture organization had a dramatic effect upon responses requiring information about the spatial location of objects. This finding is consistent with those of the earlier Mandler studies (Mandler & Parker 1976; Mandler & Johnson 1976) and with the results of Antes and Metzger (in press) in showing that contextual information is heavily involved in representing the location of objects within a scene.

The studies presented thus far lead to several conclusions regarding the effects of context on the perception of pictured information. When little specific object information is available (i.e., when pictures are viewed for a very brief time or during the initial viewing of a picture for a prolonged time) the information derived from the context of a scene is the primary basis for an observer's response to questions concerning the scene's content. If an unusual or unexpected object is placed in the scene its recognition under these circumstances is unlikely, especially if it is similar in size and shape to objects normally occupying that position in the scene. With prolonged viewing, the increasing store of detail information is relied upon to answer questions about specific object information, and with this as the basis for a response, unusual objects are quickly recognized. Viewing time does not affect the impact of contextual information as it affects responses to questions regarding the location of objects within the scene to this extent. Both initially and with prolonged viewing, context is the foundation for a spatial representation of the location of objects embedded in a pictured scene. As the contextual information available to the observer also increases with increased exposure, this spatial representation becomes more accurate with time but context remains its major informational source.

In a somewhat different approach to the study of context, Friedman (in press) has invoked the notion of "frames" to account for the processing of scene information. Noting that context must be considered to encompass both internal (memorial) and external (physical) constraints, she argued that picture perception is most probably a "top-down" affair (i.e., acquisition of information from the environment in accordance with some internally held conceptualization or hypothesis) when the object of the process is familiar or expected in relation to its context, and that it is the result of an interaction of top-down and "bottom-up" (i.e., the acquisition of environmental information in accordance with the constraints inherent in the environment itself without prejudice from any conceptualization held by the observer) processing when the object is unfamiliar or unexpected. (See Lindsay & Norman, 1977 for a discussion of data- and conceptually-driven processing in perception.) Further, she proposed that objects may be classified with respect to a given context. An "obligatory" object is one that is closely related to the theme or meaning of a scene (e.g., a refrigerator in a kitchen scene), and it functions to activate or instantiate the frame that represents that scene (in this instance, a kitchen frame). A "nonobligatory" object is one that is not directly related to the meaning of a scene (e.g., a pot of geraniums in a kitchen), but is not atypical or unexpected in that context. Nonobligatory objects are not sufficient to instantiate a scene's frame representation.

Both obligatory and nonobligatory objects are properly classified as arguments (informational units) for a given frame, and both types of objects may be inferred or activated by default in the absence of sense data to the contrary. Friedman proposed that because obligatory objects are diagnostic regarding the frames for which they are an argument, they should be processed in a top-down fashion. Nonobligatory objects will reflect a more interactive processing because they are not definitive with respect to the scene of which they are a part. Differentiating obligatory and nonobligatory objects from "unexpected" ones, she suggested that the latter represents a "wierd" element and is processed by being attached to, but not integrated with, the scene's frame. Typical objects yield object frames that easily coalesce to form larger, more global frame systems which represent the scene, whereas existing unexpected or atypical objects, and their corresponding object frames, remain differentiated. Consequently, the level of detail available to an observer will be far greater for objects that are unexpected and whose frames are not arguments of the scene frame. Theoretically then, transformations on objects (e.g., Mandler & Johnson

1976) should have different effects on recognition performance depending upon whether they are made on obligatory, nonobligatory, or unexpected objects.

Friedman (in press) tested this hypothesis by presenting subjects with six embellished (25-35 objects per picture) line drawings of real-world scenes for 30 sec each. The topic or theme of each picture was announced verbally to the subject prior to its presentation, and eye movements were recorded while subjects scanned the scenes. After viewing the entire stimulus set, subjects were given a two-alternative forced-choice recogntiion test. Every object in each of the six pictures was rated independently for its likelihood of occurrence within the picture, with high, medium, and low ratings corresponding to obligatory, nonobligatory, and unexpected objects, respectively. Type changes (i.e., replacing an object with another of the same size and shape but conceptually different), token changes (i.e., replacing an object with an object of the same conceptual class but different in size and shape), deleting an object, and exchanging the position of two objects within a picture were the four transformations carried out upon objects in the pictures used as distractors in the recognition test. Additionally, the type changes were made on objects that were rated either high, medium, or low in terms of likelihood of occurrence.

Analysis of mean fixation duration (which Friedman assumed corresponded to the time taken to encode an object) across subjects revealed that 28% to 52% of the variation was accounted for by an object's rated probability of occurrence, when considering first fixation. Likelihood accounted for between 14% and 38% on the second fixation,

1976) should have different effects on recognition performance depending upon whether they are made on obligatory, nonobligatory, or unexpected objects.

Friedman (in press) tested this hypothesis by presenting subjects with six embellished (25-35 objects per picture) line drawings of real-world scenes for 30 sec each. The topic or theme of each picture was announced verbally to the subject prior to its presentation, and eye movements were recorded while subjects scanned the scenes. After viewing the entire stimulus set, subjects were given a two-alternative forced-choice recogntiion test. Every object in each of the six pictures was rated independently for its likelihood of occurrence within the picture, with high, medium, and low ratings corresponding to obligatory, nonobligatory, and unexpected objects, respectively. Type changes (i.e., replacing an object with another of the same size and shape but conceptually different), token changes (i.e., replacing an object with an object of the same conceptual class but different in size and shape), deleting an object, and exchanging the position of two objects within a picture were the four transformations carried out upon objects in the pictures used as distractors in the recognition test. Additionally, the type changes were made on objects that were rated either high, medium, or low in terms of likelihood of occurrence.

Analysis of mean fixation duration (which Friedman assumed corresponded to the time taken to encode an object) across subjects revealed that 28% to 52% of the variation was accounted for by an object's rated probability of occurrence, when considering first fixation. Likelihood accounted for between 14% and 38% on the second fixation, and between 6% and 24% on subsequent fixations (pooled). Examining first, second, and third fixations as they interacted with likelihood, Friedman found that high and medium objects were consistently viewed for shorter times than low objects, with the difference decreasing from 342 msec on the first fixation to 78 msec on the third. Analysis of the proportion of times a distractor was correctly rejected given that the object transformed had been fixated showed that distractors were correctly rejected most often when the change involved an object with a low probability of occurrence, least often when the rated object was highly probable, with objects rated medium in likelihood intermediate between the other two. It was also found that the three type transformations and the deletion transformation were recognized a significant proportion of the time, while token and object exchange transformations were poorly detected and were within the range of chance. Finally, the effects of transformation type and likelihood of occurrence were apparently independent.

Granting Friedman's assumption that there is a correspondence between fixation duration and rate of encoding, the finding that the duration of the first fixation was a function of the a priori likelihood that the fixated object would be present in that particular scene is strong support for the hypothesis that obligatory, nonobligatory, and unexpected objects may be differentiated according to the degree of bottom-up processing evoked. As predicted, subjects rarely noticed changes to expected objects but almost always noted transformations to unexpected ones. This would indicate that the frame or schematic representation of the scene, including any important variations on its theme, was the basis for the subjects' recognition responses. Friedman asserted that providing the subject with knowledge of the general context of what would be viewed via a verbal prompt allowed the thus invoked frame structure to aid in the detection of expected objects and resulted in shorter processing times for these objects. Noting that unexpected objects are more difficult to comprehend (i.e., harder and more time consuming to encode perceptually) but easier to remember because of their differentiated representation, she concluded that the recognition of a pictured object involves the process of remembering typical or prototypical instances, afforded by the instantiation of a frame for the scene in which the object is embedded, and that confusions occur (resulting in poor recognition performance) when the task employs distractors that correspond to that frame's default information.

The research reviewed in this section provide clear evidence that contextual information is influential in the processes of perceiving and responding to pictorial information. Further, these studies demonstrate, in accordance with the duplex theory of picture perception, that the effects of this information are most pronounced when the picture is viewed for a relatively brief period. With prolonged viewing, the increase in detail or specific object information makes it probable that the nodes or arguments of the scene's schematic representation will become occupied by current input, with the result that any recognition judgment made at this time will be based upon actual rather than default information. A helpful distinction in understanding the role of contextual information in picture perception is that offered by Friedman (in press) between context as pre-existing knowledge structures and context as a set of environmental constraints present in the stimulus picture. The current investigation pursues this distinction, and the resultant effects, further.

Overview and Statement of the Problem

For the purpose of the present investigation, it is argued that the tenets of the duplex theory of picture processing, as proposed by Biederman and supplemented and amended by subsequent theory and investigation, are essentially correct. An individual possesses a set of stored knowledge structures, termed schemata or frames, which contain prototypical information about a given object or scene based on past experience. The arguments of object frames represent attribute information, and their particular structural arrangement captures the unique relations among attributes that collectively combine to result in the identity of that object. Object frames are themselves arguments of more global scene frames, and the network composed by their given architecture is synonymous with the identity of the given scene.

Typically, when an individual observes a particular scene these a priori structures are at work before the initial input is received. This occurs because the current position of the observer in the visual world is known to him or her, and this knowledge activates a related set of scene frames and their corresponding expectations. This set of a priori schemata or frames is what is meant by internal context. As information is extracted from the physical surround of the object array via peripheral systems, its processing enriches and completes the schema for that scene. The extraction and processing of specific information concerning objects from the foveal system also functions to enhance these global structures, but primarily by the direct enhancement of object frames. Thus, an observer enters a situation with a schema activated via knowledge of the visual world, and extracts contextual information from the environment to further differentiate a particular scene. The contextual information extracted from the physical environment of the object array may be referred to as external context. The schema for the scene serves to guide the extraction and interpretation of specific information gathered from the object array of the scene.

Because both internal and external sources of contextual information function to activate scene frames, it seems a reasonable assumption that their effects are additive. That is, when a scene's frame is activated prior to viewing the scene and is then embellished via information from the environmental context of the scene, subsequent object information should be extracted and processed more eficiently and accurately than when either one source of contextual information is present in isolation. Recognition of objects interpreted from this dual contextual base should be best under conditions where the alternative choices for recognition are objects which are arguments for dissimilar scene frames. When alternatives are arguments of the same scene as the target, enhancement of the recognition process should still occur, but in lesser degree, as the subject must spend less time processing the meaningful nature of the global scene, and therefore may more fully and rapidly process information about the particular arguments of the scene. The effects of the joint operation of both context sources should be least pronounced when alternatives in a recogntion test represent variations in attribute frame arguments.

The current study investigated the relationship between internal and external contexts as they interact with the type of information encoded from real world pictures. Subjects were shown line drawings of arrays of objects for 150 msec. These objects were presented either without background context or embedded in the context of a scene appropriate to those objects. Immediately prior to viewing the objects, subjects were visually exposed to the words "NO PHRASE HERE" (representing the no internal context condition), a two or three word label corresponding to the topic or meaning of the scene (representing the appropriate internal context condition), or to a label inappropriate to the topic of that scene (representing the inappropriate internal context condition). Thus, scene frames were activated prior to the stimulus presentation in two conditions, but in one that frame was inappropriate to the scene presented. A four alternative forced-choice recognition test was administered after each stimulus was presented. The three distractor items for a given set of stimulus objects were dissimilar to the target objects in either scene, object, or attribute information. A basic assumption of this study, and one supported by previous research (e.g., Loftus & Mackworth 1978), is that the utilization of contextual information in the recognition of objects is automatic, and not under subject control.

It was hypothesized that subjects would be most accurate in the object recognition task when both internal and external sources of contextual information were available, providing that information was appropriate to the scene from which the objects would typically come. When one type of contextual information was not available, it was expected that subjects would perform less well, and when available contextual information was either conflicting or inappropriate, it was expected that subjects would perform at chance levels only. Regarding the types of information varied in the response alternatives and their effects given the proposed contextual manipulations, it was hypothesized that variations in scene and object information would show the greatest effects, and that the manipulation of attribute information would be independent of contextual manipulations.

CHAPTER II

METHODOLOGY

Design

The problem under investigation required that three variables receive attention: the physical context within which stimulus objects were presented, the type of information varied in the distractor alternatives, and the existence and compatibility (relative to the physical context) of a pre-stimulus prompt. Physical context was either present or absent. The alternatives from which the target object was selected differed from the target object along one of three dimensions: (a) alternatives were objects from dissimilar scenes, (b) alternatives were different objects but from similar scenes, and (c) alternatives were the same objects but differing in some physical attribute. Context and type of information varied in the distractor alternatives were held constant across trials for each subject and varied between subjects. The pre-stimulus prompt was either compatible or incompatible with the context surrounding the stimulus object or, in a third condition, absent. Prompting conditions were varied across the trials given each subject in the experiment, i.e., prompting was treated as a within-subjects variable. Thus, the study employed a 2 X 3 X 3 factorial design with repeated measures on the third factor such that six independent groups were necessary. Accuracy of object recognition judgments in a four alternative forced-choice recognition test and confidence ratings corresponding to those judgments were used as measures of the effects of variable manipulations.

Subjects

The subjects were 72 (21 male) University of North Dakota college students participating in the experiment for credit in an undergraduate psychology course. Participation was limited to those individuals reporting normal vision without glasses or whose vision was corrected by contact lenses. Twelve subjects were randomly assigned to each of the six groups resulting from the factorial combination of the two context conditions and the three distractor conditions.

Stimuli

Pictures

The 36 pictures used as stimuli were pen and ink line drawings made by a local artist and depicted both indoor (24 pictures) and outdoor (12 pictures) still-life scenes. The artist constructed the scenes to represent 36 unique and varied themes provided by the experimenter. Care was taken to avoid the presence of animals or people, and shadowing was minimized so that the value (i.e., relative lightness or darkness) of objects was roughly equivalent. Pictured objects were typical (i.e., not unexpected--see Friedman, in press) of the scene in which they were embedded as determined jointly by the artist and the experimenter. The no-context pictures were generated by selecting six objects from each scene which represented a wide spatial distribution, reproducing them in the same locations as in the context pictures, and eliminating all other information. The resulting arrays of objects, and the context pictures, were then mounted for tachistoscopic presentation. Figures 1 and 2 illustrate a stimulus presented under the two context conditions. When projected, all pictures subtended a visual angle of 20 degrees horizontally and vertically. The objects ranged in size from 3 to 5 degrees of visual angle.

The pictures were each divided into four equal-sized quadrants and, for the 36 pictures, each of the four quadrants was randomly sampled nine times to provide the 36 target objects used in the recognition test. When more than one object was located in the sampled quadrant (this happened in only three instances) a flip of a coin determined which object would be targeted from that quadrant and for that picture.

Prompts

Two one to three word phrases were generated as descriptors of each stimulus picture, one compatible with the context of the scene and one incompatible. For the set of compatible phrases, a group of 24 undergraduate volunteers drawn from the same pool as the subjects used in the study proper were asked to give three one to three word phrases which they felt accurately summarized or described each of the 36 scenes. The most frequent descriptor from their responses was adopted as the compatible prompt for each picture. The incompatible prompt for each scene was generated by taking the less frequent responses from these subjects and pairing them with a picture expressing un unrelated theme. When the no-prompt condition was in effect, the words "NO PHRASE HERE" preceded the stimulus picture. The three types of phrases for each picture were typed in uppercase letters and centered on separate sheets of 22 X 28 cm paper for presentation.

Fig. 1. Example of targeted object (toaster in lower left-hand corner) presented under the context-present condition.

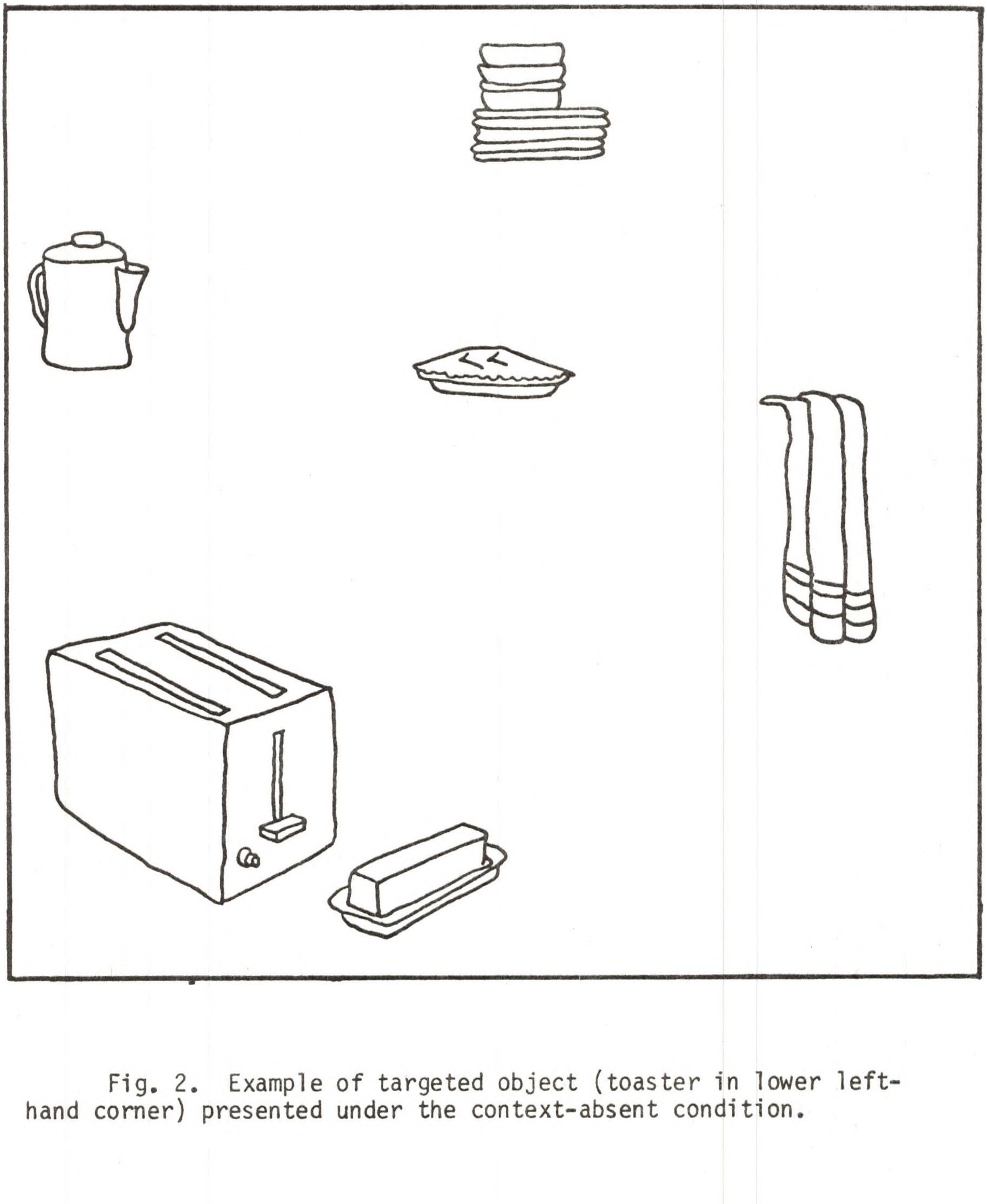

Fig. 2. Example of targeted object (toaster in lower left-hand corner) presented under the context-absent condition.

Written prompts were chosen rather than pictorial ones for two reasons. The purpose of the prompt was to activate in the subject a mental holistic representation of the scene described by that prompt. And while both written and pictorial prompts are comparable in providing holistic or semantic information about pictures under brief exposure conditions (see Potter 1975), the additional specific or detailed information available in a pictorial format might give rise to a more embellished representation than desired in the present study. The second reason was a purely economic one; written prompts are less expensive and time consuming to produce than pictorial ones.

Each subject was presented with 12 pictures preceded by compatible prompts, 12 preceded by incompatible prompts, and 12 preceded by the phrase "NO PHRASE HERE." To control for the effect of prompt presentation order, 12 different prompt orders were randomly generated with the restriction that no one type of prompt could precede more than three pictures in a row. Thus, the 12 subjects in each independent group each received a different prompt order, with the first subject in one group receiving the same order as the first subject in the other five groups, etc. In addition, the pairing of prompt types and pictures was counterbalanced, such that each picture was viewed preceded by each of the three types of prompts by four of the 12 subjects within each of the six independent groups created by the factorial combination of the two context conditions and the three distractor conditions.

Response Alternatives

The 108 response alternatives were pen and ink representations of objects drawn by the same artist who constructed the stimuli.

Alternatives varying in scene information were drawn such that they were approximately the same true size as the targeted objects but from vastly different contexts (see Figure 3). Alternatives varying in object information were different from the target object but from similar scenes (see Figure 4). The alternatives which were varied in the attribute information they contained were objects possessing the same general meaning but different along some physical dimension (see Figure 5). The objects selected as transformations of object and attribute information also approximated their respective target objects in true size.

For a given response alternative condition, the target object and three distractors were drawn in the four cells created by a 2 X 2 matrix. Each quadrant was 10 X 10 cm and the target objects were the same size in both the picture and response matrix presentations. Across the 36 pictures, each quadrant in the response matrix was occupied by a target object equally often. The four objects for each of the three response alternative conditions corresponding to each of the 36 pictures were drawn on separate sheets of 22 X 28 cm paper for presentation.

Procedure

Prior to the arrival of a subject, the experimenter determined the appropriate prompt presentation order and distractor condition and made up a looseleaf notebook that alternated prompts and response alternatives for all 36 pictures. Thus, the first page in this notebook contained the prompt for the first picture (the type determined by the presentation order in effect), followed by the response matrix

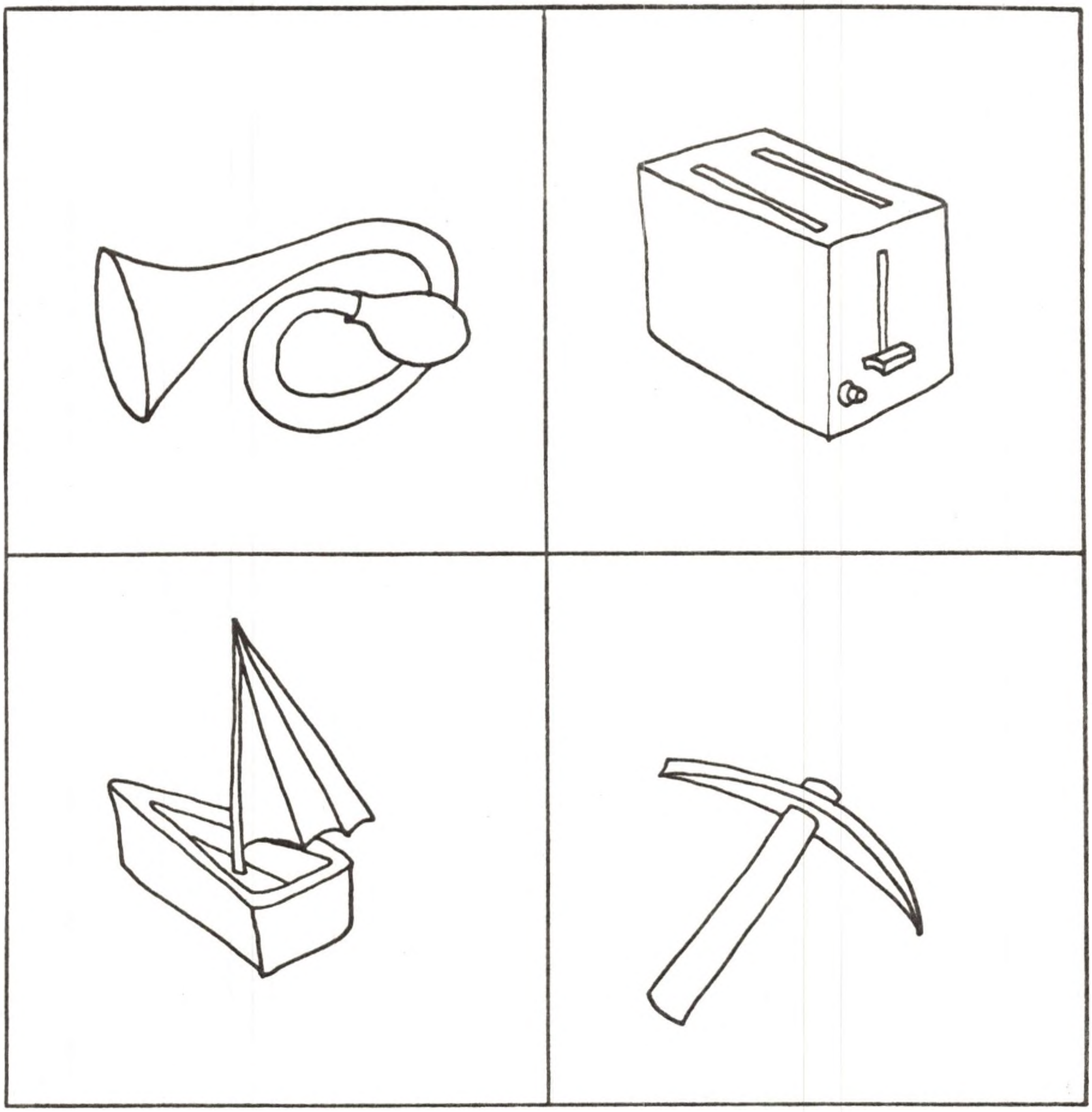

Fig. 3. Example of the response matrix employed in the forced-choice object recognition test under the scene-distractor condition. This particular matrix corresponds to the stimulus presented in Figures 1 and 2, where the toaster in the upper right-hand cell represents the targeted object and the correct choice.

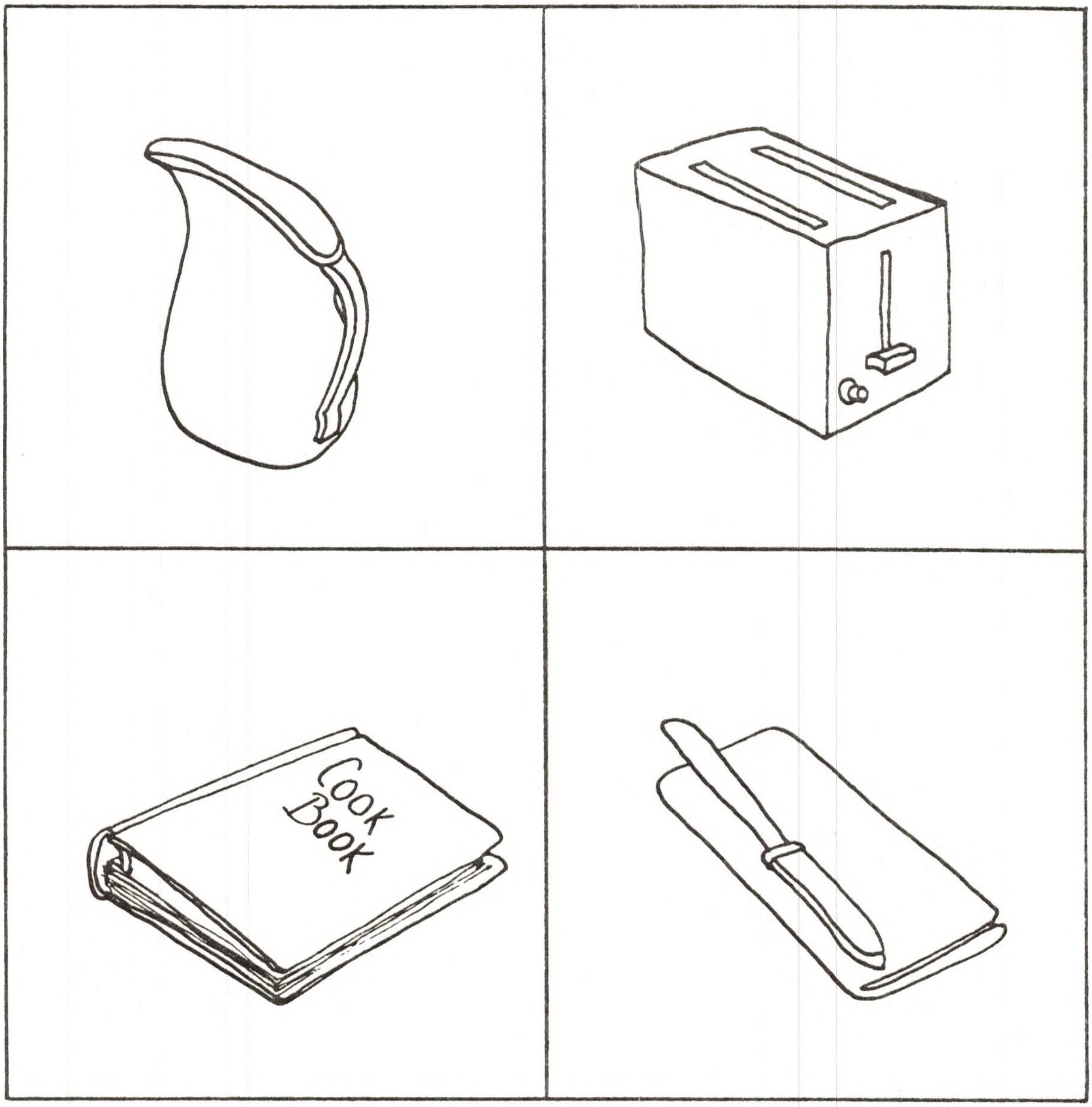

Fig. 4. Example of the response matrix employed in the forced-choice object recognition test under the object-distractor condition. This particular matrix corresponds to the stimulus presented in Figures 1 and 2, where the toaster in the upper right-hand cell represents the targeted object and the correct choice.

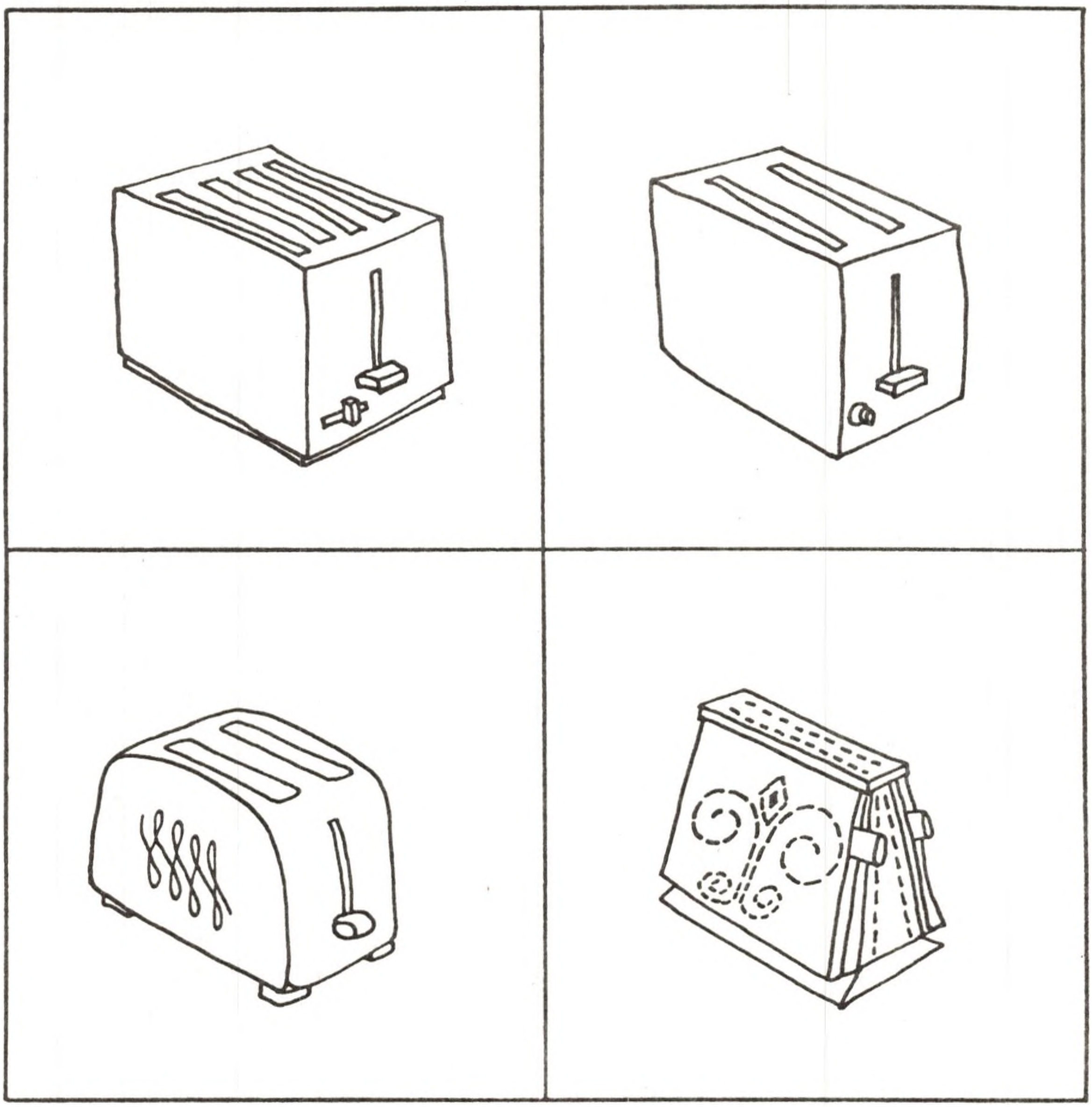

Fig. 5. Example of the response matrix employed in the forced-choice object recognition test under the attribute-distractor condition. This particular matrix corresponds to the stimulus presented in Figures 1 and 2, where the toaster in the upper right-hand cell represents the targeted object and the correct choice.

for the first picture on the second page (determined by the distractor condition in effect), followed by the third page containing the prompt corresponding to the second stimulus picture, etc. The subjects were run individually. Upon arrival, subjects were seated at a table containing the notebook and a prepared answer sheet on which to record their recognition judgments. Also located on the table was a two-channel Gerbrands Harvard tachistoscope. Prior to the first trial, subjects were read a prepared set of instructions (see Appendix A) which informed them that the study was concerned with the perception of briefly presented pictures and instructed them in the procedure of the experimental task. Further, they were told that three types of phrases were possible for any given picture, and that over the course of viewing all 36 pictures the three types would occur equally often. They were simply told to "attend" to the phrases in any case. Subjects initiated a trial by turning the page in the looseleaf notebook and silently reading the prompt for the upcoming stimulus. They then immediately looked into the tachistoscope and, as previously instructed, fixated an "X" centrally located on the screen where the picture was shown. Having indicated that they had done this by saying "READY," the stimulus picture was exposed for 150 msec. When the picture disappeared subjects looked back to the notebook where the matrix containing the target object and three alternatives was then visible (the experimenter turned the page in the notebook while the subject viewed the stimulus picture). When the decision was made as to which of the four objects in the matrix they had seen in the picture, subjects recorded that decision on the answer sheet (see Appendix B). In addition, subjects were asked to assign a confidence rating to their judgments. The ratings were made on a scale anchored at five points from "0" (representing not at all certain) through "4" (representing extremely certain), and were recorded on the prepared answer sheet. Turning the page with the response matrix for the picture just viewed exposed the prompt corresponding to the next picture and served as the beginning of the next trial. Six practice trials were succeeded by 30 experimental trials. The entire procedure, including task instructions and debriefing, required approximately 45 minutes. The luminances of the pre- and post-exposure fields were equated and approximated that of the experimental room. All trials were conducted in the same 7 X 12 ft room free of noise and occupied only by the equipment necessary for the experiment.

Data Analysis

Each subject yielded 36 responses on both the recognition accuracy and confidence measures. Only data from the 30 experimental trials were included in the analyses. Both accuracy and confidence measures were summed within the prompting conditions to obtain three totals for both measures for each subject, and these were summed across subjects within the same context X distractor condition to yield 18 sets of totals for both measures.

Data for recognition accuracy were subjected to a three factor analysis of variance with one repeated factor as an omnibus test for significance. A fixed model was assumed. Internal comparisons computed on prompting condition and distractor alternatives were accomplished using the Newman-Keuls procedure outlined in Winer (1971).

Confidence ratings for the 18 data sets generated by the factorial combination of the three variables were also subjected to an analysis of variance (fixed model); internal comparisons being computed with the Newman-Keuls procedure. While these data do not strictly conform to the scalar characteristics assumed in the employment of the analysis of variance, Tukey (1962) has empirically demonstrated that the relationship between the assignment of numbers to ordered classes and the unknown, ideal assignment along an interval or ratio scale is substantial (between .60 and .98). Thus, the utilization of the more efficient and powerful (compared to non-parametric alternatives) analysis of variance and parametric post hoc comparison techniques appears justified.

CHAPTER III

RESULTS

Recognition Accuracy

The means and standard deviations for the accuracy measure are presented in Table 1 for each of the Context-Distractor-Prompt conditions. It can be seen that the mean number of correct recognition judgments ranged from 3.583 when the target objects were viewed within the context of a scene (context-present condition), the distractors in the recognition test were conceptually equivalent to the target object but varied along some physical dimension (attribute-distractor condition), and the pre-stimulus prompt was compatible with the context of the scene in which the target object was viewed or the scene in which it is typically found (compatible-prompt condition), to 8.333 under the same context and prompt conditions but where response alternatives were objects from a scene dissimilar to that in which the target was presented or in which it is typically found (scene-distractor condition). This may be compared to a possible range for any condition of from 0.0 to 10.0, where a mean accuracy score of 2.5 would correspond to chance-level responding. Also presented in Table 1 are the means for each condition of the three major variables, collapsing across the conditions of the other two variables.

These accuracy scores were initially analyzed using a three-way Analysis of Variance applied to the Context by Distractor by Prompt factorial combination. The results of this analysis are summarized in

Table 1

Means and Standard Deviations:

Number of Correct Recognition Responses

	Prompt	Distractor: Scene	Object	Attribute	Context	Prompt
Present	Compatible	8.333 (1.303)	5.167 (1.528)	3.583 (1.311)		5.847**
	Neutral	6.833 (1.528)	4.667 (1.969)	3.667 (1.155)	5.333	5.278
Context	Incompatible	6.917 (1.832)	5.000 (1.206)	3.833 (1.193)		5.083
	Compatible	8.167 (2.038)	5.417 (1.782)	4.417 (1.782)		
Absent	Neutral	6.750 (1.288)	5.417 (1.564)	4.333 (1.614)	5.472	
	Incompatible	5.500 (2.067)	4.583 (1.443)	4.667 (1.303)		
	Distractor	7.083*	5.042	4.083		

*Mean differs significantly from the lowest mean ($p < .01$) and the next lower mean ($p < .05$)

**Mean differs significantly ($p < .01$) from the two lower means

Table 2. Both the type of distractor objects employed in the object recognition test and the compatability (relative to the physical context in which the target object was embedded or in which it is usually found) of the prompt preceding the stimulus presentation were statistically significant sources of variation. In addition, the interaction of these two variables was significant.

Using the Newman-Keuls procedure for comparing the differences between means, it was determined that target objects were recognized significantly more often when the scene-distractor condition was in effect than when distractor objects were conceptually different objects from similar scenes (object-distractor condition) or when the attribute-distractor condition was administered (see Table 1 for significance levels). Recognition accuracy under these latter two conditions did not differ significantly. Regarding the differences in accuracy for the various prompting conditions, the compatible-prompt condition resulted in significantly better performance than either the neutral-prompt (i.e., the absence of a meaningful scene description) or the incompatible-prompt (i.e., a pre-stimulus phrase which gave an inaccurate or misleading description of the stimulus scene or of the scene in which the target object would most probably be found) conditions (see Table 1 for probabilities). The observed differences between the neutral and incompatible prompting conditions failed to reach statistical significance.

Subsequent analysis of the Distractor by Prompt interaction using a test for simple effects (see Kirk, 1968) revealed that prompting manipulations had a significant effect <u>only</u> when the response alternatives in the test were objects from dissimilar scenes ($p. < .01$). The observed superiority of accuracy performance under the scene distractor condition

Table 2

Context by Distractor by Prompt ANOVA Summary:

Number of Correct Recognition Responses

Source	df	Sum of Squares	Mean Squares	F
Context (C)	1	1.04	1.04	.28
Distractor (D)	2	338.08	168.04	46.21***
C X D	2	16.08	8.04	2.20
Error (between)	66	241.42	3.66	-
Prompt (P)	2	22.69	11.35	5.96**
C X P	2	6.19	3.10	1.63
D X P	4	34.39	8.60	4.52**
C X D X P	4	4.72	1.18	.62
Error (within)	132	251.33	1.90	-
Total	215	915.95		

**$p < .01$

***$p < .001$

was significant, regardless of the particular prompting condition employed ($p < .01$). The interactive nature of these two variables is represented pictorially in Figure 6. As shown in that figure, providing subjects with an accurate description of the scene portrayed in the stimulus picture increased accuracy in the object recognition test, but only when the response alternatives in that test were objects from dissimilar scenes. When alternatives were the same object conceptually as the target but different in physical appearance or when they were conceptually different objects but from similar scenes, performance was not altered significantly. Figure 6 also illustrates the finding that scene distractors resulted in superior performance within each of the three prompting conditions, when compared to object or attribute distractors.

An unexpected finding was the failure to observe a significant interaction between the Context and Distractor factors (see Table 2). Previous research (e.g., Antes & Metzger, in press) has established that the presence or absence of scene context interacts with the type of distractor objects employed in subsequent recognition tests in affecting a subject's ability to accurately recognize objects. Antes and Metzger (in press) found that objects viewed within the context of a scene were better recognized when the response alternatives in the recognition test were objects from dissimilar scenes, whereas no differences in recognition performance were found when context-present versus context-absent conditions were compared when the response alternatives were objects that were conceptually different but from the same or similar scenes (e.g., a toaster served as an alternative to a targeted blender). As a consequence of the failure to discover this expected interaction between

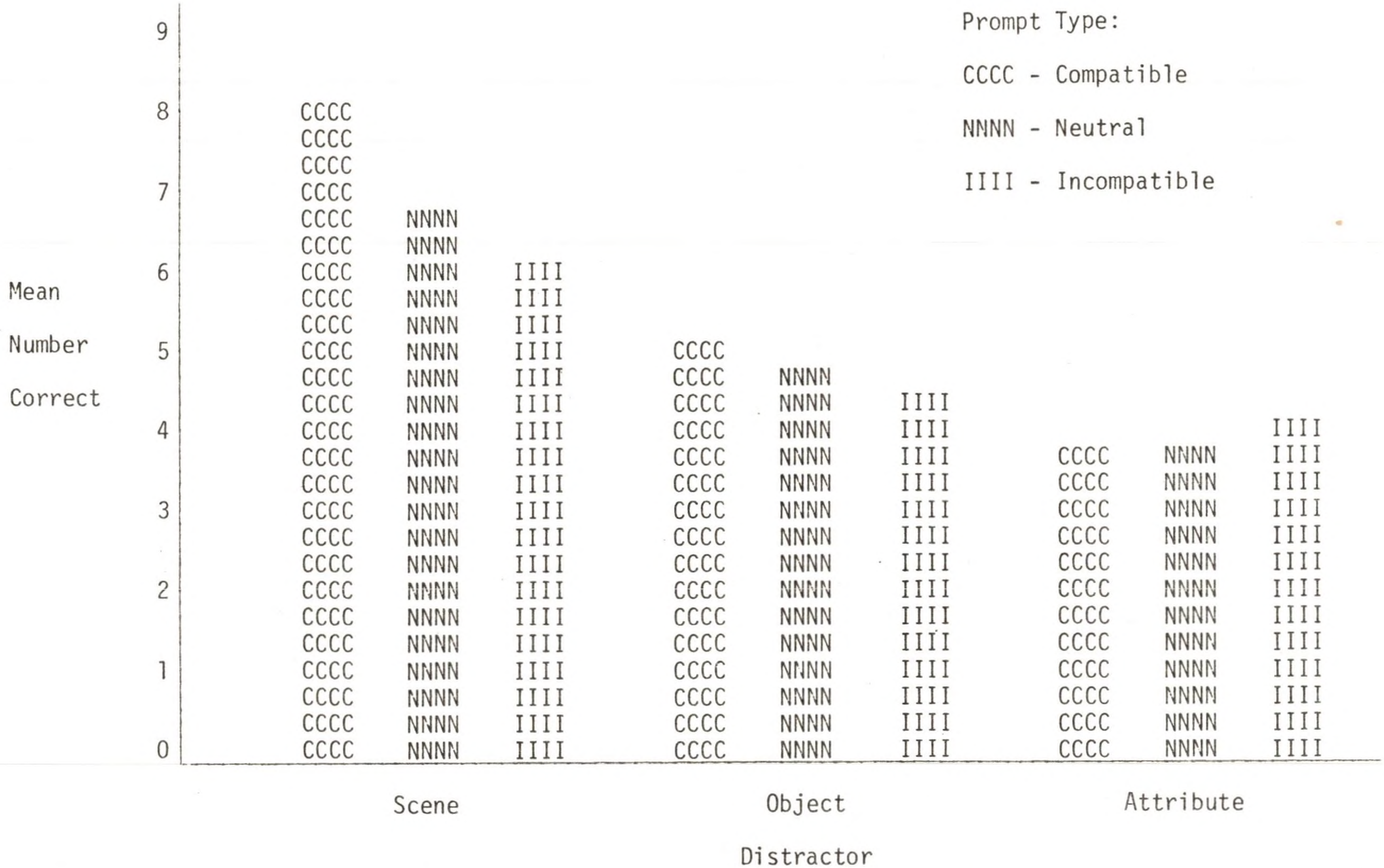

Figure 6. Mean number of correct recognition responses for the conditions resulting from the factorial combination of Distractor and Prompt variables.

Context and Distractor, additional analyses were performed in an attempt to understand this discrepancy.

An examination of the means (Table 1) for the scene and attribute distractor conditions together with those from the context-present and context-absent conditions reveals an apparent pattern. When distractors were from dissimilar scenes accuracy of recognition judgments was superior in the context-present condition, but when distractors were objects conceptually identical but with some physical attribute altered the superiority in accuracy was in favor of the context-absent condition. Differences between the two context conditions appeared nil under the object-distractor condition. Such an interrelationship between Context and Distractor may not have been revealed in the original analysis which included all three conditions of the Distractor factor, so a second Analysis of Variance was performed with the object-distractor condition omitted.

This analysis, summarized in Appendix C, resulted in the expected significant interaction between Context and Distractor ($p < .05$). A subsequent analysis of the simple effects showed that accuracy performance was best when distractors were from dissimilar scenes, regardless of the context condition ($p < .01$). The superiority in accuracy under the context-present versus context-absent conditions was near ($.05 < p < .10$) but failed to reach significance when distractor objects were from dissimilar scenes. Context-absent performance was better than context-present performance, but again not significantly ($.05 < p < .10$), when the attribute-distractor condition was in effect. Thus, while significance was not attained, a trend in the expected direction was clearly evidenced.

A more exact replication of prior researches into the combined effects of scene context and the type of distractor objects used in the recognition task (e.g., Antes & Metzger, in press) was the factorial combination of the two context conditions with the scene-distractor and object-distractor conditions employed in the present study. An Analysis of Variance performed on this combination and reported in Appendix C revealed that the Context by Distractor interaction approached, but again did not reach significant levels.

A final analysis undertaken on the accuracy measures examined the failure to achieve a significant Prompt by Context interaction (see Table 2). Such an interaction had been predicted prior to data collection, and an inspection of the means presented in Table 1 for the various conditions involved reveals that context-absent performance was better when stimuli were preceded by compatible prompts, whereas under the context-present condition there were apparently no substantial differences in accuracy between the compatible-prompt and incompatible-prompt conditions. For this final analysis then, the neutral-prompt condition was excluded and the remaining two prompt conditions were factorially combined with the two context conditions and subjected to an Analysis of Variance. Results from this analysis, summarized in Appendix C, were consistent with those obtained from the initial analysis in showing a non-significant interaction between Context and Prompt. The effects of these two variables upon recognition accuracy were apparently independent, with compatible prompts resulting in superior accuracy when compared to neutral or incompatible prompts regardless of the presence or absence of a physical context when presenting the target objects.

Confidence Ratings

Table 3 presents the means and standard deviations for the confidence ratings assigned by subjects to their recognition judgments. The possible range was from 0.0 (complete lack of confidence) to 4.0 (extreme certainty), and the obtained range for the mean confidence ratings was from 1.075 in the context-present, attribute-distractor, incompatible-prompt condition, to 2.175 in the context-present, scene distractor, compatible-prompt condition.

These confidence ratings were analyzed using a three-way Analysis of Variance applied to the Context by Distractor by Prompt combination. The results of this analysis are summarized in Table 4. Consistent with the accuracy scores, the type of distractor employed in the object recognition test and the type of prompt presented prior to the stimulus significantly affected subjects' confidence ratings. Unlike accuracy, these ratings were not differentially affected by an interaction of Distractor and Prompt, suggesting the independence of the two in their effects upon a subject's assessment of his or her success at detecting the target object.

Using the Newman-Keuls procedure for comparing means, it was found that subjects were more confident in their recognition judgments under the scene-distractor and object-distractor conditions than under the attribute-distractor condition ($p < .01$). Subjects were thus least confident when the distractor objects were conceptually identical to the target object but varied in physical appearance. No statistically significant difference was found when scene and object distractors were compared with each other, though subjects were

Table 3

Means and Standard Deviations:

Confidence Ratings

	Prompt	Distractor: Scene	Distractor: Object	Distractor: Attribute	Context	Prompt
Context: Present	Compatible	2.175 (.928)	1.833 (.549)	1.225 (.742)	1.675	1.745**
	Neutral	1.917 (.861)	1.942 (.540)	1.142 (.576)		1.635
	Incompatible	2.067 (.617)	1.700 (.598)	1.075 (.705)		1.582
Context: Absent	Compatible	1.942 (.530)	1.825 (.515)	1.475 (.533)	1.622	
	Neutral	1.800 (.433)	1.717 (.730)	1.292 (.552)		
	Incompatible	1.633 (.854)	1.733 (.576)	1.283 (.691)		
	Distractor	1.922*	1.792*	1.249		

*Mean differs significantly ($p < .01$) from the lowest mean

**Mean differs significantly from the lowest mean ($p < .01$) and the next lower mean ($p < .05$)

Table 4

Context by Distractor by Prompt ANOVA Summary:

Confidence Ratings

Source	df	Sum of Squares	Mean Squares	F
Context (C)	1	.12	.12	.13
Distractor (D)	2	18.54	9.27	9.66***
C X D	2	1.97	.99	1.03
Error (between)	66	63.37	.96	-
Prompt (P)	2	1.12	.56	3.45*
C X P	2	.03	.02	.09
D X P	4	.19	.05	.29
C X D X P	4	.51	.13	.79
Error (within)	132	21.43	.16	-
Total	215	107.29		

*$p < .05$

***$p < .001$

somewhat more confident in their responses when distractors were varied in scene information (see Table 3).

Regarding prompt effects, subjects were significantly more confident in their recognition responses when stimuli were preceded by compatible scene descriptions than when preceded by either neutral ($p < .05$) or incompatible ($p < .01$) descriptors. Subject confidence did not vary significantly as a function of whether stimuli were preceded by neutral or incompatible prompts.

Weighted Confidence Scores

In addition to accuracy and confidence scores, a third measure was created to function as an estimate of subjects' ability to appropriately place their confidence. To create such a measure, confidence ratings were signed or weighted according to whether or not the recognition response given by the subject had been accurate or inaccurate (see Palmer 1975b). Thus, a subject who gave a confidence rating of (2) to a correct recognition judgment would receive a weighted confidence score of (+2), while a subject that assigned a confidence rating of (2) to an incorrect recognition judgment received a weighted confidence score of (-2) for that stimulus. When subjects indicated that they were not at all confident in the selection that they had made (i.e., assigned a confidence rating of (0) to their recognition judgment), then they received a weighted confidence score of (0) regardless of whether their recognition response was accurate or inaccurate.

Table 5 presents the means and standard deviations for these weighted confidence scores for each of the Context-Distractor-Prompt

Table 5

Means and Standard Deviations:

Weighted Confidence Scores

Context	Prompt	Distractor: Scene	Distractor: Object	Distractor: Attribute	Context	Prompt
Present	Compatible	1.975 (.952)	.392 (.765)	-.092 (.464)	.656	.812
	Neutral	1.433 (.873)	.308 (.918)	-.008 (.392)		.621
	Incompatible	1.367 (.934)	.533 (.505)	0.000 (.407)		.621
Absent	Compatible	1.558 (.850)	.750 (.713)	.292 (.775)	.713	
	Neutral	1.233 (.818)	.600 (.441)	.158 (.691)		
	Incompatible	.925 (.689)	.533 (.576)	.367 (.440)		
	Distractor	1.415**	.519*	.119		

*Mean differs significantly (p < .01) from the next lower mean

**Mean differs significantly (p < .01) from the two lower means combinations. Here the possible range of mean score values is from -4.0 (inappropriately placing extreme confidence in incorrect recognition responses) to +4.0 (appropriately placing extreme confidence in correct recognition responses), with the value of chance or random responding undetermined. The observed range, as indicated in Table 5, was from -.092 for the context-present, attribute-distractor, compatible-prompt condition, to 1.975 for the same context and prompt conditions but where response alternatives were from dissimilar scenes.

These weighted confidence scores, representing a synthesis of the accuracy and confidence measures, were initially analyzed with a three-way Analysis of Variance applied to the Context by Distractor by Prompt factorial combination. The results of this analysis are presented in Table 6. The effect of varying the type of distractors in the recognition test significantly affected scores on this measure, as did the combination of Context and Distractor variables. The comparison of mean weighted confidence scores for the three types of distractors showed that subjects were more capable of accurately judging their performance when the test distractors were objects from dissimilar scenes than when they were either different objects from similar scenes ($p < .01$) or the same objects whose physical attributes were varied ($p < .01$). Furthermore, different objects from similar scenes resulted in a greater accuracy in confidence assignments than did the condition where objects in the recognition test differed in attribute information ($p < .01$).

The Context by Distractor interaction effects upon this weighted confidence score are depicted in Figure 7. A subsequent analysis of

Table 6

Context by Distractor by Prompt ANOVA Summary:

Weighted Confidence Scores

Source	df	Sum of Squares	Mean Squares	F
Context (C)	1	.17	.17	.24
Distractor (D)	2	63.40	31.70	43.41***
C X D	2	4.59	2.30	3.14*
Error (between)	66	48.19	.73	-
Prompt (P)	2	1.76	.88	2.33
C X P	2	.18	.09	.24
D X P	4	3.43	.86	2.26
C X D X P	4	.64	.16	.42
Error (within)	132	49.99		
Total	215	172.36		

*$p < .05$

***$p < .001$

Figure 7. Mean weighted confidence scores for the conditions resulting from the factorial combination of Context and Distractor variables.

simple effects indicated that variations in this measure approached but did not reach statistically significant levels ($.05 < p < .10$) when context-present and context-absent conditions were compared for the situation where the response alternatives were from dissimilar scenes. There was a trend in the data indicating that scores were higher under context-present than context-absent condition when distractors were from dissimilar scenes. When the two context conditions were contrasted within the object-distractor and attribute-distractor conditions the observed variations in weighted confidence scores did not even approach significant levels. It can be seen in Figure 7 that the manipulation of recognition test distractor objects resulted in significant ($p < .01$) differences when compared at both levels of the Context variable, with scene distractor conditions resulting in higher weighted confidence scores than either of the other two distractor conditions. The differences between object and attribute distractor conditions were not significant under either context conditions.

Accuracy, Confidence Ratings, and Weighted Confidence Scores Summarized

The major results from this investigation indicate that the type of response alternatives or distractors employed in the object recognition test, and the compatability of the pre-stimulus prompt (relative to the context in which the target object was viewed or the context of the scene in which it is typically found) substantially influenced both subjects' accuracy in the recognition test and their confidence in their selections. Response alternatives from dissimilar scenes consistently resulted in better recognition performance and greater confidence than when the test required that target objects be selected from among alternatives from similar scenes but conceptually different or alternatives that were conceptually identical but varied along some physical dimension. Subject confidence was a more sensitive measure of this variation, as ratings were significantly smaller (i.e., subjects were less confident) when distractors were attribute rather than object distractors. The synthesized weighted confidence score was the most sensitive measure of the three, with all three distractor types resulting in significantly different scores when contrasted with each other. Both the accuracy of the recognition judgments and subjects' confidence were enhanced when the pre-stimulus prompt was compatible with the stimulus picture (array). For neither measure were any differences significant between response given under the neutral and incompatible prompt conditions. The ability of subjects to appropriately place their confidence in an object selection was not differentially influenced by any one of the three prompting conditions.

Subject accuracy on the recognition test was also influenced by the interaction of the Distractor and Prompt variables, with greater accuracy resulting when compatible prompts preceded stimuli if the distractor objects used in the test were from dissimilar scenes, but not when distractors were from similar scenes or differed from the target object solely in physical appearance. Prompt and Distractor factors were independent in their effects upon subjects' confidence ratings and upon the subjects' accuracy in assigning those confidence ratings.

The apparent but non-significant tendency for Context and Distractor to interact evidenced in the accuracy and confidence scores reached significant levels using the weighted confidence scores.

Subsequent analyses revealed that distractor type was a substantial factor at both levels of Context, but that the presence of absence of context approached significance only for the scene and attribute distractor conditions. Looking at the Context by Distractor interaction using the recognition accuracy measure when the object-distractor condition had been excluded also resulted in overall significance, and the pattern found with the weighted confidence scores using all three distractor conditions was replicated . . . scene distractors resulted in significantly higher performance (accuracy) compared to attribute distractors, regardless of the context condition, and the presence or absence of context varied substantially but not significantly when contrasted separately under the scene and attribute distractor conditions.

CHAPTER IV

DISCUSSION

It might be helpful to begin this discussion by reviewing the theoretical foundation for this investigation as proposed at the conclusion of the Introduction. In keeping with the duplex theory of picture perception, it was proposed that one's perception and subsequent recognition of pictured objects and scenes results from the acquisition of several sources of information. One source of information is the product of a direct foveal inspection of objects within a scene such that the detail of the inspected objects is determined and available for further use by the observer. A second source of information is derived from the physical context depicted in the scene (if any), which functions as a meaningful background for the pictured objects, and in the present study is referred to as external context. The third source of information about objects and scenes resides within the observer in the form of a memorial network, organized on the basis of past experiences with similar objects and scenes, which provides information very similar to that gleaned from the scene's physical context, and in the present study is referred to as internal context. These latter two sources of information combine to result in a holistic characterization or schema of the scene, containing a more general than specific type of information, and helping to make possible an efficient and rapid interpretation of pictured objects and scenes. The detailed information about pictured objects is eventually assimilated into this representation, making it more complete and veridical. The present study was designed to allow an assessment of the separate and combined effects of these three sources of information as they function to present the observer with a meaningful visual experience of pictorially represented objects and scenes.

Recognition Accuracy

A review of past research in object and scene perception led directly to the formulation of certain hypotheses regarding the relationship of these informational sources as they might influence accuracy performance on an object recognition task. Specifically, it was expected that subjects would be most accurate on such a task when both internal and external sources of contextual information were available, provided that such information was appropriate to the scene from which the targeted objects would typically come. When one type of contextual information was not present, it was anticipated that subjects would perform less well, and when the available contextual information was either absent or conflicting, it was predicted that subjects would perform at chance levels on the recognition task. It was also hypothesized that the type of response alternatives employed in the recognition task, varied to obtain an indication of the type of information encoded and available to the observer, would interact with the presence or absence of the two sources of contextual information. The expected form of this interaction was that context manipulations would have their greatest impact on recognition accuracy when the distractors were objects from dissimilar scenes (i.e., in the scene-distractor condition), or when they were objects from similar scenes but possessing a different generic name (i.e., under the object-distractor condition). When the distractor objects were identical in name but different in appearance (i.e., the attribute-distractor condition) compared to the target object, it was anticipated that the manipulations in contextual information would be ineffectual in producing differential accuracy rates in the object recognition task.

The results obtained from this investigation were predominantly but not wholly in accord with these a priori expectations. The variation in the type of distractor objects employed in the recognition test resulted in significant differences in performance regardless of the presence or absence of contextual information (either internal or external). Further, the type of internal contextual information made available to the subject seemed to be a significant factor in its effect upon recognition accuracy, but only under the condition where the task employed distractor objects from dissimilar scenes; the expected interaction between distractor type and the type of internal context available was thus evidenced. Inconsistent with a priori predictions was the finding that the two sources of contextual information were apparently independent, with variations in internal context significantly affecting performance in the manner indicated above, while the presence or absence of a physical context failed to differentially affect performance on the recognition task. Also unexpected was the failure to find a significant interaction between the presence or absence of external context and the type of distractor objects employed in the recognition task. As noted in the Results, this interaction was indeed indicated by the data, but failed to reach statistical significance.

In the current study, subjects were essentially confronted with a discrimination task . . . accurate recognition responses were recorded when subjects were able to discriminate between a targeted object from a previously viewed stimulus and three alternative objects not previously viewed. A pertinent question is, "How difficult was the discrimination required of the subject?" The results indicated that subjects were most accurate in selecting the target object when general information about the scene from which the objects would typically come was required (i.e., the response alternatives were objects from scenes dissimilar to those in which the target object would usually be found). When more specific information concerning the identity or appearance of the target object was required, subject accuracy was significantly reduced. This latter situation occurred when distractor objects in the recognition test were objects from similar scenes but possessing a different name than the target, or when the distractors had the same generic name but were different in appearance from the target object. Empirically then, discrimination difficulty was greatest when the form of the recognition task was such that specific information about the target object was required, and least when it required only general information. This consistent difference in difficulty, found across the other two factors (internal and external context) and regardless of the interaction between the type of distractor employed and the type of internal context available, is consistent with and predictable from the duplex theory of picture perception. As discussed in the Introduction, that theory postulates that while both general and specific information about pictured objects and scenes is immediately available to the observer, the proportion of all potentially available general information exceeds the proportion of all potentially available specific information during the first few seconds of viewing. Therefore, general information is more influential in its effects upon a subject's recognition response when that response is based upon information gathered from very brief stimulus exposures, as in the present study.

The finding that recognition accuracy was superior when internal contextual information (i.e., scene descriptions or prompts) was provided prior to viewing the stimuli is complicated by the interaction of this factor with the form of the recognition task employed with a given subject (i.e., what type of distractor objects were involved). Subjects were most accurate when scene descriptions were compatible with the scene or array of objects they were to view, less accurate when the stimulus was preceded by a neutral phrase, and least accurate when the scene description was incompatible with the scene or object array it preceded. This pattern of differences was significant <u>only</u> under the testing condition where the response alternatives used in the recognition test were objects from dissimilar scenes. Such a result is highly supportive of the view that scene descriptions presented prior to the stimuli were being used primarily at the time of response (i.e., when the subject was required to select one object from among four possible objects as having been viewed previously). Under such an interpretation, internal contextual information may be utilized by the observer to rule-out or ignore objects in the recognition test that seem inappropriate to the context of the scene in which the target object was presented (or in which it is typically found). Such a response strategy, if employed by subjects in the present study, would lead to exactly the pattern of differences observed. Subjects could use the general information about the scene and objects contained in the verbal description when the form of the recognition task given the subject required only general information to make an accurate recognition judgment. Such a response strategy would be of little use when the recognition task required specific information about the targeted object. In the present study the scene descriptions were far too general (e.g., "KITCHEN SCENE") to provide the necessary specific information, and performance under such test conditions (i.e., when specific information was required) was indeed not affected differentially by the various types of prompts or scene descriptions preceding the stimulus presentation. It is expected that a similar pattern of findings would result even if _no_ pictures were presented.

The above response interpretation of the interaction between internal context and distractor type is entirely consistent with the suggestion by Antes, Penland, and Metzger (manuscript in preparation) that contextual information is utilized both during perception, and at the time of response. According to Antes, et al., available physical context influences the probability that any given pictured object will be attended and processed perceptually, and influences the course and outcome of that processing. Further, at the time of response in a recognition task, the subject goes through a two-stage decision process: The subject questions, "Do I remember seeing any of these objects in the scene (array) that I just viewed?" If the response is affirmative, then the subject selects the object that was recognized. If the response is negative, then a second question is posed, "Which of the objects available for selection most probably came from the scene (array) that I just viewed?" The subject's selection reflects the answer to this last question. The interpretation given the Distractor by Prompt interaction found in the present investigation is obviously consistent with such a two-stage response process, extended to include the influence of internal contextual information; that is, information available prior to viewing the scene (array of objects).

If varying the type (i.e., compatability relative to the scene or object array serving as the stimulus) of internal context made available to the subject in the form of a general scene description had resulted in differential accuracy performance under the attribute-distractor condition, then such a finding would evidence the perceptual effects of internally based contextual information. However, such evidence was not found in the present study. It is possible to speculate about the reasons that the proposed perceptual effects were not observed by considering the present role of prompting. Providing subjects with the description of a scene before they viewed the stimulus scene (array) represented an attempt to activate a memory network in the observer corresponding to the prompted scene . . . a network filled with general and prototypical information about the described scene, built from the products of previous perceptual experiences with such scenes. Such activation, it is presumed, is ordinarily accomplished by the observer's knowledge of where he or she is in the world at a given moment in time. That knowledge activates a set of expectations (founded upon information contained in the memorial network) about what scenes and objects the observer is most likely to encounter next. For example, it is probable that, when entering a house, one will encounter something commonly called a living room, a kitchen, bedrooms, etc., and in those rooms will observe a sofa, a toaster, and beds, etc. The current usage of verbal scene descriptions to artificially activate such a set of expectations may well have been unsuccessful. A contributing factor to this failure may have been the chosen procedure of exposing subjects to a combination of compatible, neutral, and incompatible descriptions. Knowledge that the verbal descriptions were sometimes inappropriate to the stimulus may have caused subjects to adopt a wait-and-see strategy, whereby the stimulus scene (array) was processed perceptually and the outcome of that processing was used to compare against the verbal description. If such a comparison revealed that the description provided was accurate (i.e., compatible with the scene or array in which the target object was embedded), then the description was employed at the time of response in the manner proposed in the preceding paragraph.

The failure to influence the perceptual process by prompting in the present study was also evidenced in a recent work by Biederman (1980, in press). In the study reported by Biederman, subjects given advance information (in the form of verbal scene descriptions) about the meaning of a stimulus scene were no more accurate or faster in detecting objects within those scenes or objects that had undergone violations (Biederman 1977), than subjects who had received no advance information. He argued that the process employed to perceive and interpret pictured objects and scenes is so rapid and efficient that little is gained by providing subjects with a verbal description of a scene prior to their experience with it. Consistent with the interpretation given the present results is Biederman's suggestion that what is influenced by such advance information is not perception, "but some inferential process that follows or proceeds in advance of, perception" (p. 35, 1980, in press [emphasis added]).

In that same discussion, Biederman also noted that the process of presenting subjects with a verbal description of scenes may actually interfere with the normal perceptual processes entailed in viewing and processing scene information. Ordinarily, perceptual processing of a real-world scene is not preceded by a period of verbally processing semantically related material. If such interference occurred in the present study as a result of verbally prompting subjects, it may help to explain the two unexpected findings noted earlier and discussed below.

Regarding the apparent independence of the two sources of contextual information in their effects upon recognition accuracy, it is possible that the manipulations employed in this study were simply not effective in demonstrating the impact on perception of either internal or external contextual information. The verbal prompts might have been more effective had they all been compatible in nature, but the current design (where compatible, neutral, and incompatible prompts were mixed) may have induced subjects to attempt to ignore all contextual information provided them, at least until such time as they could judge it to be accurate relative to the stimulus scene or array. An alternative possibility is that an independence of internal and external contexts truly exists, contrary to the duplex theory (see Introduction). Perhaps each type of context adds something unique to the scene characterization, such that when one source of contextual information is absent the void cannot be filled by the presence of the other. Also implied in this latter notion is that, under certain circumstances, neither source of context may be utilized in the construction of the scene characterization.

A verbal processing interference effect may also bear on the failure of the current study to replicate the findings of previous researches (e.g., Antes & Metzger, in press) showing that the presence or absence of external contextual information interacts with the type of response alternatives in the recognition test, in influencing subject accuracy. As mentioned before, such an interaction was indicated but narrowly failed to reach statistical significance. A potential explanation for this failure is that the verbal processing required of subjects interfered with the usual course of visually processing scene and object information. By design, the current investigation was supposed to be capable of testing this possibility by the inclusion of a no-prompt condition. However, as this condition still entailed a certain amount of verbal prior to visual processing (i.e., the subject still had to read the words, "NO PHRASE HERE"), it provided no clear test of this interpretation.

The strong tendency for variation of externally based contextual information to interact with the type of distractor objects employed in the recognition has important implications which justify its discussion. This finding is perhaps best explained by invoking the distinction between the top-down and bottom-up processing of pictorial information for later recognition (see Friedman, in press; Lindsay & Norman 1977; Palmer 1975b). According to this distinction, two types of information processing can be identified. One type is referred to as bottom-up or data-driven processing, indicated when processing begins with the arrival of data from the stimulus, and where each stage in the processing sequence acts upon the outcome of lower processing stages until a recognition judgment is reached. The second type of information processing is called top-down or conceptually-driven, and is characterized by processing initiated by some conceptualization or hypothesis about the stimulus to be recognized, and proceeds by seeking evidence (data) to confirm or disconfirm this initial conceptualization. If confirming evidence is available and there is little data to deny the hypothesis, that conceptualization is retained by the observer and becomes the basis of the recognition judgment. If the evidence refutes the currently operating hypothesis, a new conceptualization is adopted and the process continues. As pointed out by Lindsay and Norman (1977), both types of information processing usually occur together; that is, information is extracted and processed from the stimulus, and a conceptualization guides this processing and partially influences what information will be extracted from the stimulus in the future (i.e., with additional viewing of the stimulus). However, depending upon the conditions present at the time of viewing, one type of information processing may dominate the perceptual processing of the stimulus at a particular moment in time.

If it is realized that the recognition task in the present study functions not only as a test, but also as an effective stimulus for the subject, then it may reasonably be argued that the particular form of the test employed (i.e., the particular type of response alternatives used) stimulated subjects to adopt either a predominantly bottom-up or top-down processing strategy. The adoption of such a strategy could then be expected to result in differential accuracy performance relative to the presence or absence of physical context surrounding the object to be recognized.

As discussed earlier, general, specific, or a combination of the two types of information may be required of a subject depending upon the type of distractor objects employed in the object recognition test. General information about the scene from which the targeted object would typically come was required for an accurate recognition judgment to be made when the distractor objects were from scenes dissimilar to those in which the target was usually found. When the response alternatives in the recognition test were objects usually found in a scene similar to that in which the target was presented but having a different identity (e.g., when a toaster must be chosen from among a pitcher, cookbook, and place setting not present in the stimulus), more specific information was required for an accurate selection to occur. One might guess that in this latter situation a subject would probably process the content of the scene to the point that a list of object names was attained. In the most demanding condition in this study, the distractor objects used to assess recognition accuracy were objects that carried the same name or label as the targeted object, but had been varied along some physical dimension such that they differed from the target object in appearance. Clearly in this condition subjects would need to process the particular objects which composed the scene in considerable detail, if an accurate judgment was to ensue. Since it is a reasonable assumption that subjects would adopt the most efficient processing strategy available, it is probable that when general information was required subjects adopted a predominantly top-down approach to processing information from the scene (array), when specific information was required they utilized a predominantly bottom-up processing strategy, and when a combination of general and specific information was necessitated subjects employed both types of information processing approaches to maximize the rapid availability of both types of information.

The proposal that the distractor type evokes a particular processing approach from the subject is consistent with the interpretation given the distractor type effect presented earlier. It was noted at that time that the proportion of all potentially available general information exceeds the proportion of all potentially available specific information during the first few seconds of viewing. It is suggested that, since the top-down strategy is best suited to extracting general information and the bottom-up strategy is best suited to extracting specific information, where the observer is confronted with an extremely limited amount of viewing time, the amount of general information extracted when employing a top-down approach will exceed the amount of specific information extracted while employing a bottom-up processing approach. Thus, assuming the subject has employed the most efficient information processing strategy available to him or her, it would be expected (and it was found) that a subject faced with distractor objects from dissimilar scenes would be more accurate in recognizing target objects than one confronted with distractors that varied from the target only in physical appearance.

Support for the contention that different processing approaches are adopted as a function of varying the distractors in the recognition test may also be found in subject comments. A clear pattern was found showing that subjects in the context-present condition reported that it was much easier to determine a mismatch between the scene description (i.e., internal context) and the stimulus than to determine a match between the two. And, subjects in the context-absent condition reported that it was easier to identify a match (i.e., correspondence) than a mismatch.

Finally, the current proposal is also consistent with the suggestion by Friedman (in press) and Friedman and Bourne (1976) that the unique structural arrangement of information contained within a stimulus largely determines the depth and direction of the processing of that information. The present argument is simply that the recognition task functions as a stimulus and therefore can play an effective part in determining the processing strategy.

The question thus becomes, "What is the effect of adopting a particular processing strategy upon the availability of external contextual information as the two jointly influence recognition accuracy?" The results from the present study indicate that the adoption of a top-down processing strategy is facilitated by the presence of external context, whereas the adoption of a bottom-up approach is most effective when external contextual information is absent. That accuracy was enhanced when contextual information was present and top-down processing was used points to the character of such processing . . . contextual information is general in nature and top-down processing is founded upon general information. Further, this finding confirms what has previously been established (e.g., Antes & Metzger, in press; Biederman, et al. 1974), that physical context aids in the rapid development of a holistic characterization of the scene. When context is absent, general information is for the most part lacking, and the condition where the utilization of a top-down strategy is most probable (i.e., when distractor objects are from dissimilar scenes) yields poorer recognition accuracy. The finding that accuracy was better when a bottom-up strategy was dominant (i.e., when distractors differed from the target in appearance only) and external context was absent, when compared to the condition where such contextual information was present, requires a slightly more complex explanation.

If the processing of information from the scene (array) is bottom-up, examining the detail of individual objects and then the relations among objects, etc., then why are the context-present and context-absent conditions not equal in their effects upon subject accuracy? The answer may be three-fold: The presence of external context may result in some form of lateral inhibition or physical confusion whereby objects and other detail surrounding the target interfere with the perceptual processing of specific detail from the target, causing subjects in this condition to be less accurate. A second possibility is that there may simply be fewer objects to process in the context-absent condition such that the proportion of all available detail processed is greater in this condition than in the context-present condition, with the result that accuracy is best in the former condition. Or, the processing of general information made possible by the availability of contextual information in the context-present condition may have interfered with the encoding and storage of specific object detail. This last possibility stems from the assumption of the duplex theory that external contextual information helps to instantiate a holistic internal representation of the scene, and that a major function of such a representation is to guide attention and the course of processing to confirm expectations consistent with that representation. Subjects may have utilized the detailed object information gathered from the bottom-up processing to confirm the expectations generated by the presence of contextual information, and once confirmed, the information stored by the subject for use in the recognition test was more prototypical than specific in form. For such an interpretation to be acceptable, it must be conceded that there exists a tendency to store information in its prototypical rather than specific form whenever a more general or global scene characterization is available (see Friedman, in press), and that the presence of contextual information leads necessarily to the formulation of an integrated schematic representation of the scene from which the context was processed (see Biederman 1980, in press). According to this interpretation, recognition accuracy in the present investigation would be superior when external contextual information was absent and detail gathered via a bottom-up strategy could be encoded and remembered without interference in its original specific form. Poorer accuracy would result when the preferred bottom-up processing was interfered with by the availability of additional irrelevant (relative to the task) detail, or when interference occurred during encoding and storage.

Confidence Ratings

The present investigation also examined subjects' confidence in their recognition judgments as it varied with manipulations in contextual information and the form of the recognition test employed. While no specific a priori hypotheses concerning this measure had been proposed, it was expected that subjects would be most confident in their recognition judgments when the recognition task was administered in its least difficult format (i.e., when distractor objects were from scenes dissimilar to those in which the target objects were typically located), and when the greatest amount of contextual information (both internal and external) was available.

The observed pattern of confidence ratings were predominantly but not fully consistent with this expectation. Subjects were more confident about their responses when the recognition test demanded general rather than specific information about the scene and targeted objects, and they were more confident when internal context was provided if it was compatible with the scene or array they subsequently viewed. Unexpected was the finding that the presence or absence of an external context did not differentially affect the confidence that subjects displayed in their recognition judgments.

The finding that confidence was greater when general rather than specific information was required may be explained in terms of task difficulty, similar to the interpretation given the accuracy results for this variable. Subjects confronted with attribute distractors in the recognition test were obviously in a more demanding situation than subjects required to discriminate the target object from scene or object distractors, and their confidence ratings reflected this difference in demand. That subject confidence did not vary substantially between the scene distractor and object distractor conditions may point to the influence of general information upon a subject's feeling of confidence. In both of these conditions, as argued earlier, general information about the object targeted and the scenes in which it is normally found is required, and top-down information processing (either predominantly or equally with bottom-up processing) is most probably evoked. Perhaps subjects tend to feel more confident when they are, at least partially, guided in their observation of the visual environment by some set of expectations regarding its content. If so, then this finding seems most reasonable.

The finding that subjects were more confident in their recognition judgments under the compatible prompt condition than under either the neutral or incompatible prompting conditions may most easily be explained by looking at where in the course of the present task these scene descriptions may have had their greatest impact on the accuracy of recognition responses. As discussed earlier, it is most likely that such verbal descriptions are subjected to a wait-and-see strategy by the subject, where they are compared to the outcome of perceptually processing the stimulus and a judgment is made regarding the compatability of the scene description and the perceived stimulus scene (array). If the description is determined to be compatible it is employed at the time the recognition response is made, and if determined to be incompatible it is ignored as best as possible. Subjects were more confident when they felt they had two pieces of information (i.e., the scene description and the target object itself) on which to base their recognition response than when they felt, as a result of their own comparative and judgmental processes, that they had only a single piece of information (the object itself) on which to base that response. The failure to find any significant differences in assigned confidence ratings when the neutral and incompatible prompting conditions were compared, supports the notion that subjects tended to ignore the incompatible scene descriptions (i.e., treat them as though they were neutral or nonexistent descriptions) when they were assessing how much confidence they were willing to place in their recognition responses.

That subject confidence did not vary substantially depending upon the presence or absence of external contextual information reinforces the suggestion made earlier that employing a mixture of compatible, neutral, and incompatible scene descriptions prior to the stimulus presentation may have interfered with the ordinary course of visually processing information from a scene (array), to the extent that the normal influence of external contextual information upon the perception of an object was disrupted. It is conceivable that such a disruption led subjects to discount the presence or absence of this type of contextual information in assigning confidence to their recognition judgments.

The lack of a Distractor by Prompt interaction using the confidence measure, vis-a-vis the presence of that interaction with the accuracy measure, may be interpreted as evidence that regardless of the difficulty level of the recognition test subjects felt more confident and secure in their responses when they had information about the context of the stimulus, and they knew (judged) that information to be compatible. As discussed earlier, accuracy was significantly better when compatible (compared to neutral or incompatible) scene descriptions were provided, but only when the difficulty of the discrimination was at a minimum.

Weighted Confidence Scores

Prior to discussing the findings with this measure, it is important to consider its nature. The weighted confidence scores represent a combination of subjects' recognition accuracy scores and their ratings of confidence associated with the recognition responses. Thus, this measure does not correspond to any overt response given by subjects and therefore must be interpreted cautiously. As the score is weighted in such a way as to result in higher values when a high confidence rating is paired with an accurate recognition response and the lowest values result when a high confidence rating has been assigned to an inaccurate recognition judgment, one interpretation of these scores is that they represent the ability of subjects to appropriately or accurately place their confidence. It is in this sense that the weighted confidence scores were employed in the present study and for this discussion. The synthesis of accuracy and confidence measures was used to determine if the discrepancy between the pattern of findings for these two measures studied in isolation could be better understood, and to discover if an integrated measure might reveal the anticipated external context by distractor type interaction indicated by the accuracy data.

The presence and compatability of a pre-stimulus verbal scene description interacted with the type of distractor objects employed in the recognition test in their effects upon subject accuracy in that test. However, subject confidence did not reveal this interaction; both the presence and compatability of a pre-stimulus scene description and the type of distractors used in the test significantly affected subject confidence, but their effects were apparently independent. When this interaction was examined using weighted confidence scores it failed to reach significance. Further, these scores were not differentially influenced by the three prompting conditions. That subjects were no better or worse at judging their recognition performance when the effects of the three types of verbal scene descriptions were compared may be interpreted by proposing that only a knowledge of task difficulty is effective in allowing subjects to accurately judge their recognition responses as correct; information concerning the nature of the scene to be viewed may be beneficial (or detrimental, if incompatible) to the recognition response and may influence how much confidence the subject places in that response, but it does not provide the subject with a good basis for appropriately placing confidence. One explanation for this finding is that this "extra" information about the nature (meaning) of the scene lulled the subjects into a false sense of security about the accuracy of their recognition judgments.

Such an interpretation is supported by the finding that scores on this measure were substantially higher when the form of the recognition test included distractor objects that were from dissimilar scenes than when the response alternatives were objects from similar scenes but different in identity, or the same objects but varied in appearance. When subjects felt they were faced with a difficult discrimination in the recognition test, they were least able to accurately judge their recognition responses. As the difficulty lessened, they became increasingly more proficient in judging these responses.

As presented in the Results, weighted confidence scores also varied differentially and significantly as a result of the interaction between the presence or absence of external context and the type of distractor objects employed in the recognition task. That subjects were more capable of accurately assigning confidence ratings to their response when external contextual information was present and the distractors were objects from dissimilar scenes, and were also more accurate when context was absent and the distractors differed from the target object in appearance only, suggests that the demand for object detail when processing from the bottom up, and the demand for general information about the object when scene information is being processed from the top down may create an ideal circumstance for subjects to be able to judge the accuracy of their own responses. When the demand is for general information about the scene and only an array of objects is available, subjects must struggle with the decision as to which processing strategy is most likely to be profitable, and that struggle inhibits an accurate assessment of performance. Likewise, when specific detail about pictured objects is required by the task at hand and additional irrelevant detail is presented (i.e., context-present, attribute-distractor condition), a similar struggle may ensue and similarly place the subject in the worst position possible for being able to accurately judge his or her own performance.

The interpretations given findings with the weighted confidence scores assume that a subject's awareness of his or her accuracy on the task employed in the present study is actually reflected in their assignment of confidence ratings to the responses given in that task. Further, it is necessary to point out that this measure is completely dependent upon the accuracy and confidence measures, with its uniqueness due to the particular combination of these measures. In consequence, the finding that certain variable manipulations significantly affected this measure when they had similar effects upon accuracy and confidence considered apart, may be expected on the basis of this dependent relationship. The potential value of this measure appears to lie in its ability to offer additional information concerning the situation which arises when the anticipated consistency between the two primary measures is not found. It may or may not be a more sensitive measure of the effects of variable manipulations than the two measures associated with overt responses.

Summary and Implications

It is possible to summarize the findings of the present investigation by considering the influences of the two types of information available in a pictorial representation of real world scenes and of the demands imposed on the subject by the particular task employed to study the perception of these representations.

It may be inferred from the duplex theory of picture perception that when pictures of scenes and objects are viewed for a very brief period of time, more general than specific information is processed from these pictures. Object recognition performance in the present study reflects the differential availability of these two kinds of information, as both accuracy and confidence were greatest under conditions where general, as opposed to specific or detailed, information about the object and scene tested was required to discriminate the targeted object from the alternatives. With increased viewing time (i.e., longer stimulus exposures), the duplex theory suggests that the difference in the availability of general and specific object information would become less pronounced, eventually reducing to zero as the processing of both general and specific information reached an asymptote. While the present study did not employ extended exposure times, it can be predicted that the increase in viewing time would result in the absence of accuracy and confidence differences for the scene and attribute distractor conditions studied here.

The duplex theory also postulates that one source of general information about a pictured scene is in the form of a memorial network, organized on the basis of past experiences that the observer has had with similar scenes, which facilitates the recognition of objects embedded in that scene. This facilitation may occur at two places during the recognition task, either during the processing and encoding of the object and/or at the time the recognition response is made. In the current study, an attempt was made to activate this memory network by providing subjects with a verbal description of a scene (e.g., "KITCHEN SCENE"). It was found that such verbal prompting did in fact help in the recognition of objects, but apparently this influence was limited to the time the response was given. It was suggested that the failure of the verbal scene descriptions to influence the course of perceptual processing in the present study was due to the employment of incompatible and neutral, as well as compatible scene descriptions. Because of this, subjects were induced to adopt a wait-and-see strategy regarding the use of these scene descriptions until such time as they could make a judgment regarding their compatability. That judgment required that the stimulus scene or array of objects be at least partially processed, and thus the manipulation had little effect upon the perception of the objects. A possible test for the perceptual effects of this internally based contextual information would be the pre-stimulus administration of compatible (i.e., appropriate) scene descriptions to one group of subjects, providing a second group of subjects with no pre-stimulus information, and then comparing the performance of the two groups on an object recognition test using distractors similar to those in the attribute-distractor condition of the present study. If, as the duplex theory proposes, providing the observer with general information prior to his or her exposure to the scene facilitates the efficient extraction of detailed or specific object information from that scene, then any differences in object recognizability between two such groups could reasonably be attributed to the differential impact of the presence or absence of internal a priori general information upon the perceptual process.

Subjects in the present study were also more confident in their recognition responses when they felt they had two (both a scene description and the object itself) rather than a single (just the object) piece of information on which to base those responses. That subjects were most confident when scene descriptions were compatible with the scene (array) they actually viewed, assigning equally low ratings to their recognition judgments when the stimulus had been preceded by either a neutral or incompatible scene description, is additional evidence that a comparative process is occurring between the information provided prior to the stimulus (i.e., the pre-stimulus prompt) and the information gathered directly from the stimulus by the subject.

Finally, the particular format of the task utilized in testing object recognizability may partially determine the type of information processing strategy engaged in by the subject. Information may be processed from a scene (array of objects) in a predominantly bottom-up fashion, in a predominantly top-down fashion, or it may be processed by using a combination of these two approaches. Depending upon the amount of specific object detail required by the recognition task for an accurate response to result, the adoption of one of these three general processing strategies will occur. It was proposed that when the task demands only general information concerning the object to be recognized, a predominantly top-down approach to processing information will be selected. When the recognition test demands a detailed knowledge of the targeted object, a predominantly bottom-up processing strategy is adopted, and when the test necessitates the acquisition of both general and specific information for an accurate discrimination, a combination of the two information processing approaches is used. In addition, it was suggested that each particular processing strategy is either facilitated or inhibited by the presence of physical context in the stimulus scene (array). According to the duplex theory, such external contextual information provides the observer with a more general than specific type of information about the pictured scene and the objects contained within it. On this basis, a reasonable conclusion is that the adoption of a predominantly top-down processing strategy is facilitated by the presence of external context, whereas the selection of a predominantly bottom-up strategy is facilitated by the absence of such contextual information.

The results from the present study are supportive of this conceptualization; subjects were most accurate in the object recognition task when its format required general information and external context was available, and when its format demanded specific detail and external contextual information was absent. However, these results reflect a trend rather than significant differences in performance. It was suggested that the failure of these differences to reach significant proportions may have been due to some degree of interference with the processing of external context by the presence of scene descriptions which were varied in appropriateness to that context. For this reason, a verification of the relationship between information processing strategies and the contextual information available to the observer, as proposed in the preceding paragraph, should be conducted with verbal scene descriptions excluded from the procedure.

Taken as a whole, the findings of this investigation are quite supportive of the tenets of the duplex theory of picture perception. In addition, they point to the ever-present role of the experimental task as an influence upon subjects' reactions and responses. It remains the task for future research to assess the nature of contextual information, and to explicate the critical elements which contribute to its formulation and activation as a major factor in our perception of pictures.

APPENDICES

APPENDIX A

INSTRUCTIONS READ TO ALL SUBJECTS

This study is concerned with how people perceive pictures. Specifically, I am interested in the kinds of information people receive when viewing a picture for a very brief time, and their ability to utilize that information in an object recognition test.

Basically, the task involves viewing a (an) picture (array of objects), presented very briefly, and then deciding which of four objects came from that picture (array). Prior to viewing each picture (array of objects), you will be required to read a short phrase written in this notebook [point]. The phrase may simply be the words "NO PHRASE HERE," or it may consist of a meaningful description of a pictured scene. On one-half of the trials, this description will be an accurate description of the picture (array of objects) that you will view . . . on the other half, the description will not be accurate. Of course you will have no way of knowing which is the case until you actually see the picture (array). In any case, it is important that you attend to the phrase you read. After you have read the phrase, you are to look into this viewing aperature [point] and stare at the "X" on the display screen. When you have located the "X" and are looking directly at it, say "READY," and shortly thereafter a (an) picture (array of objects) will be flashed on the screen. Once the picture (array) disappears, decide which one of the four objects pictured in the notebook was contained in the picture (array) you saw on the screen. Indicate your response on the answer sheet in front of you. Also, I am asking you to indicate the amount of confidence you feel you can place in each selection you make. You can do this by circling the appropriate number on that same sheet. Following this, turn the page in the notebook and the phrase corresponding to the next picture (array of objects) will be visible, and the next trial will begin.

To summarize the steps involved in the task: read the phrase in the notebook, fixate the "X" on the display screen and then say "READY" [when you look into the viewing aperature I will turn the page in the notebook with the phrase so that the four objects are visible], after the picture (array of objects) disappears look to the notebook and select the object you think was contained in the presentation, indicate this selection on the answer sheet, and circle the number that corresponds to the degree of confidence you have in that particular selection.

Do you have any questions concerning what you are supposed to do?

OK. There will be six practice trials followed by thirty test trials. At the conclusion of all thirty-six trials, I will explain in more depth the nature of the study and inform you of your performance on the object recognition test. At that time I will also try to answer any questions you may have.

APPENDIX B

ANSWER SHEET USED TO RECORD SUBJECTS' RESPONSES

Answer Sheet

Name ________________________

A	B
C	D

Write the number corresponding to the cell containing the object from the display on the blank to the left of the trial number.

0	1	2	3	4
not at all confident	.	.	.	extremely confident

Circle the number that corresponds to the amount of confidence you place in your selection.

P1. ____ 0 1 2 3 4

P2. ____ 0 1 2 3 4

P3. ____ 0 1 2 3 4

P4. ____ 0 1 2 3 4

P5. ____ 0 1 2 3 4

P6. ____ 0 1 2 3 4

1. ____ 0 1 2 3 4

2. ____ 0 1 2 3 4

3. ____ 0 1 2 3 4

4. ____ 0 1 2 3 4

5. ____ 0 1 2 3 4

6. ____ 0 1 2 3 4

7. ____ 0 1 2 3 4

8. ____ 0 1 2 3 4

9. ____ 0 1 2 3 4

10. ____ 0 1 2 3 4

11. ____ 0 1 2 3 4

12. ____ 0 1 2 3 4

13. ____ 0 1 2 3 4

14. ____ 0 1 2 3 4

15. ____ 0 1 2 3 4

16. ____ 0 1 2 3 4

17. ____ 0 1 2 3 4

18. ____ 0 1 2 3 4

19. ____ 0 1 2 3 4

20. ____ 0 1 2 3 4

21. ____ 0 1 2 3 4

22. ____ 0 1 2 3 4

23. ____ 0 1 2 3 4

24. ____ 0 1 2 3 4

25. ____ 0 1 2 3 4

26. ____ 0 1 2 3 4

27. ____ 0 1 2 3 4

28. ____ 0 1 2 3 4

29. ____ 0 1 2 3 4

30. ____ 0 1 2 3 4

APPENDIX C

SUMMARY TABLES FOR POST HOC ANOVA'S REPORTED IN TEXT

Table 7

Context by Distractor[a] by Prompt ANOVA Summary:

Number of Correct Recognition Responses

Source	df	Sum of Squares	Mean Squares	F
Context (C)	1	.34	.34	.09
Distractor (D)	1	315.06	315.06	86.27***
C X D	1	16.67	16.67	4.56*
Error (between)	44	160.70	3.65	-
Prompt (P)	2	24.67	12.33	6.36**
C X P	2	1.72	.96	.44
D X P	2	30.17	15.08	7.78***
C X D X P	2	3.56	1.78	.92
Error (within)	88	170.56	1.94	-
Total	143	723.44		

a - Object condition excluded from Distractor factor in analysis

*p < .05

**p < .01

***p < .001

Table 8

Context by Distractor[a] by Prompt ANOVA Summary:

Number of Correct Recognition Responses

Source	df	Sum of Squares	Mean Squares	F
Context (C)	1	1.36	1.36	.31
Distractor (D)	1	144.00	144.00	32.83***
C X D	1	5.44	5.44	1.24
Error (between)	44	192.94	4.38	-
Prompt (P)	2	40.62	20.31	10.42***
C X P	2	8.60	4.30	2.20
D X P	2	16.62	8.31	4.26*
C X D X P	2	.60	.30	.15
Error (within)	88	171.56	1.95	-
Total	143	581.75		

a - Attribute condition excluded from Distractor factor in analysis

*$p < .05$

***$p < .001$

Table 9

Context by Distractor by Prompt[a] ANOVA Summary:

Number of Correct Recognition Responses

Source	df	Sum of Squares	Mean Squares	F
Context (C)	1	.00	.00	-
Distractor (D)	2	247.72	123.86	40.67***
C X D	2	15.17	7.58	2.49
Error (between)	66	201.00	3.04	-
Prompt (P)	1	21.78	21.78	10.63**
C X P	1	2.78	2.78	1.36
D X P	2	29.56	14.78	7.22**
C X D X P	2	2.72	1.36	.66
Error (within)	66	135.17	2.05	-
Total	143	655.89		

a - Neutral condition excluded from Prompt factor in analysis

**$p < .01$

***$p < .001$

APPENDIX D

TABLE 10

Table 10

Means for Each Subject Used in Analysis Reported in Text

Subject # (1)	Context (2)	Distractor (3)	Compatible Prompt: Number Correct (4)	Compatible Prompt: Average Confidence (5)	Compatible Prompt: Weighted Confidence (6)	Neutral Prompt: Number Correct (7)	Neutral Prompt: Average Confidence (8)	Neutral Prompt: Weighted Confidence (9)	Incompatible Prompt: Number Correct (10)	Incompatible Prompt: Average Confidence (11)	Incompatible Prompt: Weighted Confidence (12)
12	Present	Scene	7	.5	.5	5	1.4	1.2	5	1.6	1.6
14	"	"	9	3.0	2.8	8	2.9	2.7	9	3.0	2.8
18	"	"	7	2.0	1.6	9	2.4	2.2	7	1.9	.9
19	"	"	9	2.6	2.6	8	2.8	2.8	7	2.6	2.0
36	"	"	6	1.3	.9	5	1.3	.9	8	2.2	2.2
38	"	"	9	.7	.7	6	.4	.4	7	1.6	1.6
42	"	"	9	2.8	2.6	7	.7	.5	8	2.2	2.0
43	"	"	9	2.6	2.4	6	1.4	1.0	5	.9	.3
60	"	"	10	3.5	3.5	6	2.8	.6	5	3.1	.3
62	"	"	7	1.9	1.7	5	2.0	.8	4	2.0	- .4
66	"	"	8	2.4	1.6	8	2.7	2.1	8	1.8	1.2
67	"	"	10	2.8	2.8	9	2.2	2.0	10	1.9	1.9
2	Absent	Scene	10	1.9	1.9	8	1.8	1.8	3	.3	.3
3	"	"	8	1.6	1.2	7	1.8	1.6	8	2.1	1.5
7	"	"	6	.7	.7	7	1.6	1.6	2	0	0
8	"	"	8	2.5	2.3	8	2.7	2.1	6	2.8	1.2
26	"	"	10	2.2	2.2	8	2.3	1.9	7	2.6	1.8
27	"	"	10	1.4	1.4	8	1.0	1.0	7	1.6	1.2

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
31	Absent	Scene	10	2.2	2.2	6	1.9	.9	8	2.0	2.0
32	"	"	10	2.7	2.7	6	1.8	1.6	6	1.1	.9
50	"	"	4	2.1	.1	5	1.5	- .1	4	1.6	.2
51	"	"	6	1.8	.2	4	1.6	- .6	5	2.4	.7
55	"	"	9	2.0	2.0	7	1.5	1.5	7	1.7	1.3
56	"	"	7	2.2	1.8	7	2.1	1.5	3	1.4	0
9	Present	Object	4	2.9	- .3	6	2.7	1.5	6	2.4	.6
15	"	"	4	1.1	.7	6	1.0	.8	7	1.3	1.1
16	"	"	2	1.6	-1.0	8	2.2	1.6	6	1.7	.3
17	"	"	5	1.3	- .1	5	2.1	.5	5	2.0	1.0
33	"	"	6	2.5	.9	7	2.0	1.0	6	1.9	1.1
39	"	"	7	2.2	.8	4	1.8	0	6	2.8	1.4
40	"	"	7	2.2	1.4	3	2.2	- .8	4	1.6	.4
41	"	"	7	2.2	1.2	3	2.1	- .7	5	1.3	.3
57	"	"	4	1.3	- .5	1	1.3	- .9	3	1.2	- .2
63	"	"	6	1.7	1.1	5	1.1	.7	4	.5	.3
64	"	"	5	1.5	- .1	5	2.4	.8	4	1.7	.1
65	"	"	5	2.1	.6	3	2.4	- .8	4	2.0	0
1	Absent	Object	4	1.6	.6	7	1.9	.9	2	1.7	- .5
5	"	"	4	1.3	.3	5	1.7	.5	4	1.8	.6
11	"	"	5	1.5	.1	6	1.3	.5	4	1.8	- .2
24	"	"	5	2.2	.2	6	2.8	.8	6	2.7	.3
25	"	"	6	2.3	.7	6	2.4	.8	6	2.2	1.2
29	"	"	2	1.0	- .2	1	.4	- .4	3	.9	.5
35	"	"	8	1.4	1.4	6	1.4	1.4	6	1.5	.3
48	"	"	4	2.6	.1	6	2.9	.7	4	1.8	.6
49	"	"	8	2.3	1.7	5	1.5	.3	5	1.7	.7
53	"	"	6	1.5	.7	5	1.2	.2	6	1.7	1.3

(1)	(2)	(3)	(4)	(5)	(6)	(7)	)8)	(9)	(10)	(11)	(12)
59	Absent	Object	7	2.4	2.0	5	2.0	.8	6	2.4	1.4
72	"	"	6	1.8	1.4	7	1.1	.7	3	.6	.2
4	Present	Attribute	2	1.0	- .5	5	1.6	.4	5	1.4	.8
10	"	"	4	.5	.3	5	.8	.4	3	.4	- .4
13	"	"	3	.5	.3	3	.9	.3	4	.4	- .2
22	"	"	4	1.4	.4	3	1.4	- .2	6	1.5	.5
28	"	"	3	1.7	- .3	2	1.5	- .9	5	2.0	- .4
34	"	"	3	.7	- .5	2	.3	- .1	3	.5	- .1
37	"	"	1	1.4	-1.0	4	.8	- .2	3	.8	- .3
46	"	"	4	1.1	- .3	3	.8	0	3	1.1	.1
52	"	"	4	1.5	- .1	5	1.3	- .3	5	1.2	.4
58	"	"	6	1.6	.6	3	1.6	- .2	2	1.0	.2
61	"	"	5	.2	.2	4	.4	.4	3	.1	- .1
70	"	"	4	3.0	- .2	5	2.3	.3	4	2.5	- .5
6	Absent	Attribute	5	1.9	.5	3	1.3	.1	2	1.7	- .3
20	"	"	6	2.1	.5	6	2.3	.7	6	2.3	.3
21	"	"	7	1.3	1.3	3	.7	.3	4	.2	0
23	"	"	1	1.8	-1.6	2	2.1	-1.2	4	1.8	0
30	"	"	3	1.4	- .4	5	1.6	.4	6	1.3	.7
44	"	"	3	1.5	.7	5	.8	.2	7	1.8	1.4
45	"	"	5	1.9	.1	3	.9	-1.0	4	1.3	.1
47	"	"	3	.4	0	6	.6	.6	5	.7	.5
54	"	"	4	.9	- .1	2	1.2	- .4	5	.8	.6
68	"	"	5	1.9	1.1	6	1.8	1.0	4	2.2	.6
69	"	"	4	.8	.8	6	1.0	.8	5	.4	.2
71	"	"	7	1.8	.6	5	1.2	.4	4	.9	.3

REFERENCES

Anderson, J. R. Arguments concerning representations for mental imagery. Psychological Review, 1978, 85, 249-277.

Antes, J. R. The time course of picture viewing. Journal of Experimental Psychology, 1974, 103, 62-70.

Antes, J. R. Recognizing and localizing features in brief picture presentations. Memory and Cognition, 1977, 5, 155-161.

Antes, J. R. & Metzger, R. L. Context effects on the recognition and localization of objects from pictures. Acta Psychologica, in press.

Antes, J. R., Penland, J. G. & Metzger, R. L. Context effects on the recognition of usual and unusual objects in pictures. Manuscript in preparation.

Antes, J. R., Singsaas, P. A., & Metzger, R. L. Components of pictoral informativeness. Perceptual and Motor Skills, 1978, 47, 459-464.

Antes, J. R. & Stone, L. A. Multidimensional scaling of pictorial informativeness. Perceptual and Motor Skills, 1975, 40, 887-893.

Avant, L. L. & Helson, H. Theories of perception. In B. Wolman (Ed.), Handbook of general psychology. Englewood Cliffs, New Jersey:

Bartlett, F. C. Remembering: A study in experimental and social psychology. Cambridge: Cambridge University Press, 1932.

Berlyne, D. E. Conflict, arousal, and curiosity. New York: McGraw-Hill, 1960.

Biederman, I. Perceiving real-world scenes. Science, 1972, 177, 77-79.

Biederman, I. On processing information from a glance at a scene: Some implications for a syntax and semantics of visual processing. In S. Treu (Ed.), User-oriented design of interactive graphics systems. New York: ACM, 1977.

Biederman, I. On the semantics of a glance at a scene. In M. Kubovy & J.R. Pomerantz (Eds.), Perceptual organization. Hillsdale, New Jersey: Lawrence Erlbaum, in press.

Biederman, I., Glass, A. & Stacy, E. Searching for objects in real-world scenes. Journal of Experimental Psychology, 1973, 97, 22-27.

Biederman, I., Rabinowitz, J., Glass, A., & Stacy, E. On the information extracted from a glance at a scene. Journal of Experimental Psychology, 1974, 103, 597-600.

Boring, E. G. A history of experimental psychology, 2nd edition. Englewood Cliffs, New Jersey: Prentice-Hall, 1950.

Broadbent, D. A. Decision and stress. London: Academic Press, 1971.

Bruner, J. S. Personality dynamics and the process of perceiving. In R. R. Blake & G. V. Ramsey (Eds.), Perception--An Approach to Personality. New York: Ronald Press, 1951.

Bruner, J. S. Going beyond the information given. In H. Gruber, et al. (Eds.), Contemporary approaches to cognition. Cambridge: Harvard University Press, 1957(a).

Bruner, J. S. On perceptual readiness. Psychological Review, 1957 (b), 64, 132-152.

Bruner, J. S. & Klien, G. S. The functions of perceiving: New look retrospect. In B. Kaplan & S. Wapner (Eds.), Perspectives in psychological theory: Essays in honor of Heinz Werner. New York: International Universities Press, 1960.

Brunswik, E. Perception and the representative design of psychological experiments. Berkeley: University of California Press, 1956.

Buswell, G. W. How people look at pictures. Chicago: University of Chicago Press, 1935.

Day, M. C. Developmental trends in scanning. In H. W. Reese (Ed.), Advances in child development and behavior (Vol. 10). New York: Academic Press, 1975.

Estes, W. K. & Taylor, H. A. Visual detection in relation to display size and redundancy of critical elements. Perception & Psychophysics, 1966, 1, 9-16.

Evans. S. H. A brief statement of schema theory. Psychonomic Science, 1967, 8, 87-88.

Friedman, A. Framing pictures: The role of knowledge in automatized encoding and memory for gist. Journal of Experimental Psychology: General, in press.

Friedman, A. & Bourne, L. E. Encoding levels of information in pictures and words. Journal of Experimental Psychology: General, 1976, 105, 169-190.

Garner, W. R. Uncertainty and structure as psychological concepts. New York: Wiley, 1962.

Gibson, E. J. Principles of perceptual learning and development. New York: Appleton-Century-Crofts, 1969.

Gibson, J. J. The perception of the visual world. Boston: Houghton-Mifflin, 1950.

Gibson, J. J. The senses considered as perceptual systems. Boston: Houghton-Mifflin, 1966.

Gibson, J. J. The information available in pictures. Leonardo, 1971, 4, 27-35.

Gould, J. D. Pattern recognition and eye-movement parameters. Perception & Psychophysics, 1967, 2, 399-407.

Gould, J. D. & Dill, A. D. Eye-movement parameters and pattern discrimination. Perception & Psychophysics, 1969, 6, 311-320.

Gould, J. D. & Peeples, D. R. Eye-movements during visual search and discrimination of meaningless, symbol, and object patterns. Journal of Experimental Psychology, 1970, 85, 51-55.

Gregory, R. L. The intelligent eye. New York: McGraw-Hill, 1970.

Head, H. Studies in neurology. London: Hodder & Stoughton and Oxford University Press, 1920.

Hebb, D. O. The organization of behavior. New York: Wiley, 1949.

Hochberg, J. In the mind's eye. In R. N. Haber (Ed.), Contemporary theory and research in visual perception. New York: Holt, Rinehart & Winston, 1968.

Hochberg, J. Attention, organization and consciousness. In D. I. Mostofsky (Ed.), Attention: Contemporary theory and analysis. New York: Appleton-Century-Crofts, 1970.

Hochberg, J. The representation of things and people. In E. H. Gombrich, J. Hochberg & M. Black, Art, perception, and reality. Baltimore: The Johns Hopkins University Press, 1952.

Hochberg, J. Motion pictures, eye movements and mental maps: Perception as purposive behavior. Paper presented at the meeting of the American Psychological Association, Chicago, 1975.

Hochberg, J. Perception, 2nd edition. Englewood Cliffs, New Jersey: Prentice-Hall, 1978.

Ittleson, W. H. & Cantril, H. Perception, a transactional approach. New York: Doubleday, 1954.

James, W. The principles of psychology. New York: Holt, 1890.

Kahneman, D. Attention and effort. Englewood Cliffs, New Jersey: Prentice-Hall, 1973.

Karpov, B. A., Luria, A. R. & Yarbus, A. L. Disturbances of the structure of active perception in lesions of the posterior and anterior regions of the brain. Neuropsychologica, 1968, 6, 157-166.

Kennedy, J. M. A psychology of picture perception. San Francisco: Jossey-Bass, 1974.

Kirk, R. E. Experimental design: Procedures for the behavioral sciences. Belmont, California: Brooks/Cole, 1968.

Koffka, K. Principles of Gestalt psychology. New York: Harcourt, Brace & Co., 1935.

Lindsay, P. H. & Norman, D. A. Human information processing: An introduction to psychology. New York: Academic Press, 1977.

Loftus, G. R. & Bell, S. M. Two types of information in picture memory. Journal of Experimental Psychology: Human Learning and Memory, 1975, 104, 103-113.

Loftus, G. R. & Mackworth, N. H. Cognitive determinants of fixation location during picture viewing. Journal of Experimental Psychology: Human Perception and Performance, 1978, 4, 565-572.

Mackworth, N. H. & Bruner, J. S. How adults and children search and recognize pictures. Human Development, 1970, 13, 149-177.

Mackworth, N. H. & Morandi, A. J. The gaze selects informative details within pictures. Perception & Psychophysics, 1967, 2, 547-552.

Mandler, J. & Johnson, N. Some of the thousand words a picture is worth. Journal of Experimental Psychology: Human Learning and Memory, 1976, 2, 529-540.

Mandler, J. & Parker, R. E. Memory for descriptive and spatial information in complex pictures. Journal of Experimental Psychology: Human Learning and Memory, 1976, 2, 38-48.

Mandler, J. & Ritchey, G. Long-term memory for pictures. Journal of Experimental Psychology: Human Learning and Memory, 1977, 3, 386-396.

Metzger, R. L. Processing pictorial information: Effects of informativeness, location and duration. Unpublished doctoral dissertation, University of North Dakota, 1976.

Minsky, M. A framework for representing knowledge. In P. H. Winston (Ed.), The psychology of computer vision. New York: McGraw-Hill, 1975.

Miller, G. A., Galanter, E. & Pribram, K. H. Plans and the structure of behavior. New York: Holt, Rinehart & Winston, 1960.

Neisser, U. Cognitive psychology. New York: Appleton-Century-Crofts, 1967.

Neisser, U. Cognition and reality. San Francisco: W. H. Freeman, 1976.

Northway, M. L. The concept of the schema. Parts I & II. British Journal of Psychology, 1940, 30-31, 316-325, 22-36.

Noton, D. & Stark, L. Eye movements and visual perception. Scientific American, 1971, 224, 34-43.

Oldfield, R. C. & Zangwill, O. L. Head's concept of the schema and its application in contemporary British psychology. Parts I, II, III, & IV. British Journal of Psychology, 1942-1943, 32-33, 267-286, 58-64, 113-129, 143-149.

Palmer, S. E. The effects of contextual scenes on the identification of objects. Memory and Cognition, 1975 (a), 3, 519-526.

Palmer, S. E. Visual perception and world knowledge: Notes on a model of sensory-cognitive interaction. In D. A. Norman & D. E. Rumelhart (Eds.), Explorations in cognition. San Francisco: W. H. Freeman, 1975(b).

Palmer, S. E. Hierarchical structure in perceptual representation. Cognitive Psychology, 1977, 9, 441-475.

Pastore, N. Selective history of theories of visual perception: 1650-1950. New York: Oxford University Press, 1975.

Piaget, J. The construction of reality in the child. Translated by M. Cook. New York: Basic Books, 1954.

Piaget, J. & Inhelder, B. Memory and intelligence. New York: Basic Books, 1973.

Pollack, I. & Spence, D. Subjective pictorial information and visual search. Perception & Psychophysics, 1968, 3, 41-44.

Pomerantz, J. R., Sager, L. C. & Stoever, R. J. Perception of wholes and their component parts: Some configural superiority effects. Journal of Experimental Psychology: Human Perception and Performance, 1977, 3, 422-435.

Potter, M. C. Meaning in visual search. Science, 1975, 198, 965-966.

Pylyshyn, Z. W. What the mind's eye tells the mind's brain: A critique of mental imagery. Psychological Bulletin, 1973, 80, 1-24.

Rayner, K. Eye movements in reading and information processing. Psychological Bulletin, 1978, 8, 618-660.

Reicher, G. M. Perceptual recognition as a function of meaningfulness of stimulus material. Journal of Experimental Psychology, 1969, 81, 275-280.

Sloman, A. Interactions between philosophy and artificial intelligence: The role of intuition and non-logical reasoning in intelligence. Artificial Intelligence, 1971, 2, 209-225.

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

Titchener, E. B. Lectures on the elementary psychology of feeling and attention. New York: MacMillan, 1919.

Tolman, E. C. Cognitive maps in rats and men. Psychological Review, 1948, 55, 189-208.

Tukey, J. W. Data analysis and behavioral science. Preliminary draft of manuscript. Princeton University and Bell Telephone Laboratories, 1962.

Tulving, E. & Gold, C. Stimulus information and contextual information as determinants of tachistoscopic recognition of words. Journal of Experimental Psychology, 1963, 66, 319-327.

Turvey, M. T. Perspectives in vision: Conception or perception? In D. Duane & M. Rawson (Eds.), Reading, perception and language. Baltimore: York Press, 1975.

Turvey, M. T. Contrasting orientations to the theory of visual information processing. Psychological Review, 1977, 84, 67-88.

Vernon, M. D. A further study of visual perception. Cambridge: Cambridge University Press, 1952.

Vernon, M. D. The function of schemata in perceiving. Psychological Review, 1955, 62, 180-192.

Vernon, M. D. Perception through experience. Edinburgh: T. and A. Constable Ltd., 1970.

Wachtel, P. L. Conceptions of broad and narrow attention. Psychological Bulletin, 1967, 68, 419-429.

Weisstein, N. & Harris, C. S. Visual detection of line segments: An object superiority effect. Science, 1974, 186, 752-755.

Wheeler, D. D. Processes in word recognition. Cognitive Psychology, 1970, 1, 59-85.

Williams, L. G. The effect of target specification on objects fixated during visual search. Perception & Psychophysics, 1966, 1, 315-318.

Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill, 1971.

Winston, P. H. Learning structural descriptions from examples. In P. H. Winston (Ed.), The psychology of computer vision. New York: McGraw-Hill, 1975.

Wolters, A. W. P. On conceptual thinking. British Journal of Psychology, 1933, 24, 133-143.

Woodworth, R. S. Experimental psychology. New York: Holt, 1938.

Yarbus, A. L. Eye movements and vision. Translated by Basil Haigh. New York: Plenum Press, 1967.