The identification of objects in scenes : the role of scene backgrounds on object naming.

Item Type	dissertation
Authors	Boyce, Susan J.
DOI	10.7275/1rq0-7e06
Download date	2025-03-21 00:30:44
Link to Item	https://hdl.handle.net/20.500.14394/11139

THE IDENTIFICATION OF OBJECTS IN SCENES: THE ROLE OF SCENE BACKGROUNDS ON OBJECT NAMING

A Dissertation Presented

by

SUSAN J. BOYCE

Submitted to the Graduate School of the
University of Massachusetts in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

February, 1990

Department of Psychology

THE IDENTIFICATION OF OBJECTS IN SCENES:
THE ROLE OF SCENE BACKGROUNDS ON OBJECT NAMING

A Dissertation Presented

by

SUSAN J. BOYCE

Approved as to style and content by:

Alexander Pollatsek

Alexander Pollatsek, Chair

Donald L. Fisher

Donald Fisher, Member

Katherine Fite

Katherine Fite, Member

Keith Ray

Keith Rayner, Member

Arnold Well

Arnold Well, Member

Melinda Novak

Melinda Novak, Department Chair

Department of Psychology

ACKNOWLEDGMENTS

The two people who deserve most of the thanks for their support and guidance are Sandy Pollatsek, my official advisor, and Harry Blanchard, my unofficial advisor. Working with Sandy has truly been a pleasure. His creativity and sensitivity played such an important role in my graduate career. Harry has not only served as my emotional support, but has spent countless hours in the last four years challenging my ideas and truly teaching me to think. I owe both of them a great deal for their support.

Many of my fellow graduate students have made life bearable and often times, even fun. Robin Morris has been a great friend and an excellent research partner. Through collaboration (and many lunches) with her I learned to communicate and cooperate. Sara Sereno has so many times cheered me up. She made me laugh when I didn't find life very funny. I hope to find colleagues like them again in my future.

Finally, I would like to thank the members of my dissertation committee, Keith Rayner, Arnie Well, Kay Fite and Don Fisher for their suggestions and cooperation during this project.

ABSTRACT

THE IDENTIFICATION OF OBJECTS IN SCENES: THE ROLE OF SCENE BACKGROUNDS IN OBJECT NAMING

FEBRUARY, 1990

SUSAN J. BOYCE, B.S., URSINUS COLLEGE
M.S., UNIVERSITY OF MASSACHUSETTS
Ph.D., UNIVERSITY OF MASSACHUSETTS

Directed by : Professor Alexander Pollatsek

The purpose of this research was to investigate the role of scene backgrounds on object identification. Previous research with brief presentation of scenes indicated that scene context facilitated object identification. Experiment 1 replicated this finding with longer display durations. Experiments 2 and 3 were designed to investigate the time course of background information acquisition using an eye movement paradigm. Although the results from Experiment 2 were inconclusive, Experiment 3 demonstrated that scene background information was acquired on both the first and second fixations on a scene. It was concluded that background information acquired from the first and second fixations facilitates object identification.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Understanding the process of object identification is essential for our understanding of both human visual processing capabilities and our understanding of the organization of human memory. The human ability to transform light falling on to the retina into meaningful information is an amazing capability that we accomplish effortlessly and without awareness. However, the cognitive processes involved in making sense of our visual world are complex and currently not very well understood.

On a broad level, the purpose of this dissertation is to add to our knowledge of the cognitive processes involved in human visual information processing. Object recognition is just one of the tasks our visual system accomplishes. It has been chosen for study primarily because it is a definable task. That is, operationally we can define object recognition as having a beginning (when light reflected off the object falls onto the retina) and an end (when information stored in memory about that object has been accessed).

Much of the research on object identification has focused on the issue of whether recognition is the result of a bottom-up perceptual analysis that proceeds independently of other cognitive processes or whether object recognition is highly dependent on world knowledge stored in memory to aid the perceptual processes. This is an important first step to understanding visual processing because by determining what information is used in object identification we can begin to outline how the process is carried out.

There are many different ways to investigate the process of object identification. One obvious way is to study how people recognize single objects in isolation, realizing that this is a somewhat artificial situation. Objects in the real world do not appear isolated but instead occur as part of much more complicated "scenes". However, it is thought that if we can explain the process of the identification of a single object in isolation then explaining object identification under more naturalistic settings should follow logically. That is, the basic processes involved in object recognition should be the same whether or not the object is located in a scene context.

On the other hand, the process of object identification may proceed very differently when an object is a part of a more naturalistic scene than when it is in isolation. That is, something about the scene information itself may play a role in the object identification process. If this were the case, then studying the process of object identification with isolated objects may never bring us closer to the goal of more generally understanding human visual information processing.

Overview

The rest of this chapter is devoted to a review of what we currently know about object identification and how scene context affects the process of object identification. The first section contains a review of a recent model of object recognition proposed by Biederman (1987). The model that he puts forth is a bottom-up account of object recognition that is based upon the idea that to recognize an object one must first recognize the component parts of an object. I believe this model to be the state-of-the-art in modelling the object recognition process. Much about the model is untested and, in fact, may be untestable, however, I think it is important to include because it provides some insight into where we currently are in our understanding of object recognition. The second section of this chapter deals with the literature on object recognition in scene context. The general purpose of all this research was to determine whether scene context plays a role in the identification of objects.

A Model of Object Recognition

The most recent and most comprehensive model of object recognition has been put forth by Biederman (1987). The basic idea behind his Recognition-by-Components (RBC) model is that objects are recognized by the identification of the component "parts" of the object. Biederman's model begins describing the process of object identification after a stage of edge extraction has already been completed. At the level in visual processing where his model begins, the representation is very much like a line drawing of an object. The first stage in his model is the segmentation of the object into its component parts. This segmentation is accomplished by an algorithm that does not depend on knowledge about the object's identity, but instead relies only on the information available in the edge-based representation. The resulting parts are then matched against a set of primitive shapes stored in memory. This results in a description of the object's shape in terms of this constrained set of primatives. Finally, this representation is matched against information in memory about the object's identity.

In some respects, this model is like the old feature models of letter recognition (e.g. Selfridge and Neisser, 1960). Both depend on a componential form of analysis. That is, the entire stimulus is not recognized as a whole. Instead it is broken up into features, or parts, that are identified first and this leads to the recognition of the whole.

The advantage of this kind of model over a template model is that it attempts to define in a constrained way what sort of information is stored in memory and how the matching occurs between incoming visual information and stored information in memory. Biederman (1987) claims that all objects are composed of some combination of a limited set of volumetric shapes (his name for these is geons). The identity of an object is defined by which particular volumetric primitives it is made of and the spatial relations of these components to one another. For example, a coffee mug is made up of two geons, a cylinder geon (makes the cup part) and a curved cylinder (which makes the handle).

The part of the model that is least well defined is exactly how these relations between the geons become specified. That is, a coffee mug is not the only object made up of a straight cylinder geon and a curved cylinder geon. These same geons define the shape of a bucket. Biederman and others (Rock, 1983, for example) propose that a structural description is formed that specifies the relations between the parts. Thus, the structural description for the coffee mug would differ from that of the bucket in the way in which the geons were attached together. So far, Biederman has not specified how this structural description is formed. Nor has he clearly defined whether a structural description is like a map (spatially organized) or some kind of list of propositions.

For all of its problems, Biederman's model is probably one of the most detailed models of object recognition to date. Perhaps the best aspect of his model is that he has precisely defined an algorithm that segments an object into constituent parts that only relies on the visual input. The parts that result of this segmentation algorithm correspond well with our phenomonological impression of what should be a part (dividing a coffee cup into the cup part and the handle part, for example). Furthermore, there is some behavioral evidence to suggest that when the information that serves as input to the algorithm is missing from an object, identification is extremely difficult or impossible (Biederman and Blickle, 1985). Also, his explanation of exactly how the parts are fitted by the appropriate geon is fairly detailed and depends only upon the information contained within a two-dimensional "line-drawing like" representation.

Perhaps the knottiest problem with the RBC model lies in testing the validity of these shape primitives; geons. After the object has been segmented into its constituent parts, the resulting part is compared to the set of 36 primitives. The nearest match will then be used to describe the shape of that region of the object. This "nearest match" situation makes it impossible to really test whether he has properly defined the set of visual primitives, because in his definition everything will be "fit" by one of the geons to some approximation. Thus, there really is no such thing as a part that is not a geon and this makes this aspect of the model somewhat untestable.

Overall, however, Biederman has begun to outline a framework of object recognition which is detailed enough to make a number of predictions and advance our knowledge about the object recognition process. The problems with the model are that the definition of "geon" is sufficiently vague that it may not be useful and the concept of the structural description is still left very unclear. But the bottom-up nature of this model has some appeal.

Validity of the Edge Based Approach

The point must be made that the RBC model rests on the assumption that perceptual analysis is preceded by an edge detection process (i.e. that further processing of an image proceeds on a representation that is virtually a line drawing, no matter what the input was like). Thus, information about color and texture (surface characteristics) plays no role in the initial processing of the stimulus. Information about the surface characteristics does not in any way affect the initial access to the memory representation for that object under most normal situations. It could be argued that this assumption is self-serving. That is, it is much easier to do this kind of research with line drawings as opposed to color photographs, so hence the assumption.

In fact, there does exist some data that would argue that surface characteristics cannot be ignored in models of object recognition. Ostergaard and Davidoff (1985) presented subjects with photographs of fruits and vegetables in their first experiment. The photographs were either black and white or in color and the subjects had to name the object. They found a significant advantage in object naming times when the objects were presented in color over black and white. In a second experiment they tested whether this advantage for color photographs would occur in an object recognition task, as well as a naming task. They wanted to hold color constant so they used three stimuli that had a similar color (tomato, radish and strawberry) and photographed them in color, in black and white, and then, in order to test the effect of the inappropriate color, they spray painted the objects blue and took color photographs. One group of subjects named the objects as before. For a second group of subjects, a target object was specified (the radish for example) and they were to respond yes/no on each trial as to whether it was the radish. They again found an advantage in naming time for color photographs (the correct color) over black and white. (They found no difference between the black and white trials and the blue trials). However, this advantage for correct color slides did not exist on the recognition task. Thus, they conclude that color facilitates object naming, but not object recognition.

There are aspects of this study that do not seem quite right. First of all, it could be the case that there are classes of objects in which color is important for identification. The objects chosen by Ostergaard and Davidoff seem to fall into that category. Many fruits and vegetables are similar in terms of their shapes and therefore it must be something about their surface characteristics that distinguish them. So they may have stacked the deck to get color effects by the choice of their stimuli. Secondly, I find it surprising that if color is so important, that they did not get an interference effect for objects that are the inappropriate color (blue strawberry, for example). This could be because there was such a limited set of objects (three, shown for hundreds of trials) and that the same inappropriate color was used for each object.

Biederman and Ju (1988) conducted their own study to determine whether color of an object really is important to the identification process. They presented subjects with line drawings of objects or color photographs of the same objects and recorded both naming times and recognition times as did Ostergaard and Davidoff (1985). They found no difference in either measure (naming or recognition) between the color photographs and the line drawings. They then went back and classified their objects as having a diagnostic color or not using the rationale that, perhaps for some classes of objects, color is more important. They performed an items analysis on the basis of this sorting and still found no advantage for color photographs. From this they conclude that color information is not important for the first contact with the memory representation, and that only later on in the process should color/surface information be registered.

Their diagnostic/non-diagnostic color distinction is not terribly impressive. The examples of objects with diagnostic color are banana, fork, fish and camera as opposed to chair, pen, mitten and bicycle pump for non diagnosticity of color. It seems as though there are other objects that would be better candidates for the diagnostic color category, such as orange, apple, leopard, tiger, etc. However, Biederman does concede that there may be classes of objects where the only distinguishing features between the members are their surface characteristics, and he proposes that identification of these objects would take longer.

The issue of whether the identification of line drawings is exactly the same as identification of more naturally represented objects is obviously not one that has been settled. On some level it is hard to believe that surface characteristics, such as color and texture, play no role in object identification. On the other hand, the behavioral evidence collected on this issue seems mixed. The research conducted as part of this dissertation used line drawings of objects and scenes as stimuli. The decision to use line drawings as representations of objects and scenes in the real world was made for two reasons: a) since it is not clear that the perception of photographs is qualitatively different than the perception of line drawings and b) working with line drawings allows the experimenter much more control over the stimulus display.

Biederman's RBC model, as it currently stands, accounts for the process of object recognition without calling upon world knowledge to supplement the perceptual processes. He has outlined a framework of recognition that is dependent only on the information contained within a two-dimensional representation of the visual stimulus. In the following section evidence will be reviewed that suggests that context plays an important role in object recognition. Although Biederman's RBC model currently does not indicate the role of scene context on object identification it could easily be adapted to incorporate scene context effects.

The Role of Scene Context in Object Identification

Much of the research in picture perception has focused on the issue of how context affects the object identification process. It is presumed that by studying the situations where object identification is speeded or impaired we might gain a better understanding of the processes involved in identifying objects in the natural world.

It has been repeatedly demonstrated that objects located in a coherent scene context are identified more accurately than objects located in an incoherent scene context (Antes, 1974; Antes, 1977; Antes and Mann, 1984; Antes and Penland, 1981; Antes, Penland and Metzger, 1981; Biederman, 1972; Biederman, 1981; Biederman, Mezzanotte and Rabinowitz, 1982; Biederman, Rabinowitz, Glass and Stacy, 1974; Boyce, Pollatsek and Rayner, 1989; Friedman, 1979; Loftus and Mackworth, 1978). A number of theories have been developed to explain this phenomenon. The basic idea behind these theories is that something about the scene context facilitates the initial contact with a memory representation for the target object. The two important aspects of these theories are the locus of the scene context effects (that is, what in the scene is actually doing the facilitating) and the mechanism of facilitation (how the facilitation happens). I will begin with a discussion of the literature concerning the locus of scene context effects and then deal with the issue of mechanism.

Locus of Scene Context Effects

Something present in the scene context must be responsible for the more rapid identification of objects located in coherent scenes. In order to investigate which aspects of a coherent scene are important one must know what makes a scene seem coherent. Biederman, Mezzanotte and Rabinowitz (1982) attempted to define scene coherence as the relationships between the objects in a scene. They proposed that a scene could be considered coherent if all the objects obeyed the following principles of relationships: Probability - an object is likely to occur in the context; Size - an object is in correct proportion to the rest of the scene; Position - an object is in the correct location in the scene; Support - an object appears to obey the laws of gravity and; Interposition - an opaque object occludes what is behind it.

Biederman et al. (1982) presented scenes for 150 ms and measured percent correct on an object identification task. They found that an object violating one or more of these relation principles was harder to identify than the same object without a violation. Biederman et al. suggested that scene context effects were due to these relationships between the objects. That is, early in scene viewing these relationships are identified and through the extraction of these relationships, appropriate memory structures are activated and the object identification process is facilitated.

This argument, however, is circular. If accessing this relational information facilitates one's ability to identify a target object, then it must come before the object has been identified. But all of these relations seem to be dependent on first knowing the identity of the object. That is, how can you know that a sofa does not belong in the street scene, unless you have already identified the context and, more importantly, that the object in question is a sofa and not a truck?

Biederman at al. counter that some of the relations are dependent on knowing the identity of the object (probability, position and size) while others can be accessed without knowing the meaning of the object (support, interposition). His argument is that one may know that an object is lacking support, such as a sofa floating in mid-air, from a low level parsing of the background that indicates that the sofa is not resting on a solid surface and that this parsing stage occurs very early in the processing of the scene (before you have identified the object). Parsing, in this context refers to the assignment of line segments as belonging to an object or belonging to some other entity in the scene. However, even if one can determine that an object lacks support in the scene very early in processing, one still has to know that it is a sofa to realize that it is violating a relation because some objects do not require support (e.g. airplanes). Therefore, it is not clear that these scene relations defined by Biederman et al., can be accessed prior to the identity of objects in the scene.

Loftus and Mackworth (1978) claimed that something about the global meaning of the scene or "gist" is important in facilitating the processing of the objects, using as evidence where people chose to look in a scene. They monitored eye movements while subjects viewed scenes for 4 seconds each. Some of the scenes contained objects that were very unlikely to occur given the scene context (i.e. octopus in a farm background, tractor in an underwater background). They found that low probability objects were fixated earlier, were fixated more often, and were fixated for longer durations. From these data Loftus and Mackworth claimed that subjects readily obtain the "gist" of the scene (within the first fixation) and partially identify objects in the periphery. This partial object identification then leads subjects to compute the probabilities that these objects are likely to occur given the gist and the eye is guided to the low probability objects first.

Loftus' model places a lot of importance on the first fixation on the scene. In his view, it seems as though most of the processing required to understand the scene is conducted in the first 150 ms. Even the objects are "partially" identified. This is a claim that will be addressed later in the paper. A second problem with Loftus' model is his dependence on fixation duration as a dependent measure. It could be that subjects fixated the octopus longer in the farm context because it was interesting or anomalous. That is, subjects could have identified the octopus as rapidly in the farm scene as in the underwater scene, but were having problems making sense of the scene. Loftus' results cannot distinguish between these two explanations. Furthermore, subjects might have seen their task as "finding the strange object" and this led to the target object being fixated early in scene viewing.

Boyce, Pollatsek and Rayner (1989) conducted a series of experiments in an attempt to be more diagnostic about what aspects of scene context facilitate object identification. They reasoned that a scene could be thought of as a background and a collection of objects. Objects were defined as small closed figures that subtended approximately two degrees of visual angle and had the characteristic that they were "movable" entities. That is, they were things that could be moved around in the real world (for example, a toaster). A "background" was defined as a large entity, taking up most of the visual field (approximately 15 degrees) and had the characteristic of being locally meaningless. That is, any small region of the background would not convey the meaning of the background. The visual information that contained the meaning was distributed over almost the entire visual field. Boyce et al. were interested in the role of scene backgrounds in object identification. In the first experiment objects were located either in a consistent background (bedroom objects in bedroom background), in an inconsistent background (bedroom objects in refrigerator background) or in no background (see figure 1). Subjects were first presented with a the name of a target object. The scene then appeared for 150 ms followed by a pattern mask consisting of randomly placed line segments and angles. A filled circle located within the pattern mask served as a location cue indicating the possible position of the target object. Subjects were to respond 'yes' or 'no' as to whether the target named at the beginning of the trial was present at the cued location. This paradigm was exactly the same as Biederman's (1982). The purpose of Boyce et al.'s research was to a) replicate Biederman's context effect with the violation of probability and b) to be more precise about the locus of this context effect (i.e. scene backgrounds instead of object relations).

Subjects were more accurate at identifying objects in the consistent backgrounds than in the inconsistent backgrounds. However, subjects were equally accurate in the consistent backgrounds and the no background control. Hence, context affected performance, but the effect appeared to be interference from the inconsistent backgrounds rather than facilitation from the consistent backgrounds.

The second experiment attempted to determine the relative contributions of backgrounds and non-target objects on target object identification using the same task. In Experiment 1, every scene was made up of five objects that belonged together in a scene (e.g. bedroom objects) and only the consistency of the background with this object set was manipulated. In the second experiment, both the consistency of the background (with respect to the target object) and the consistency of the cohort (non-target) objects were independently varied. The results indicated that there was no effect of cohort consistency. That is, all of the effect of context observed in the first experiment was due to the degree of background consistency and the cohort objects contributed nothing.

The third experiment was designed to determine whether the observed effects of context were due to consistent backgrounds facilitating object identification or inconsistent background interfering with object identification. In Experiment 1 performance with the consistent backgrounds was equal to performance with the no background controls. Thus it appeared as though inconsistent backgrounds interfered with object identification. However, it could have been the case that performance was relatively good in the no background controls for a different reason than that producing the good performance in the Consistent Background condition. That is, perhaps the good performance in the No Background controls was because these scenes were less visually complex than the scenes with backgrounds. In order to test this hypothesis, nonsense backgrounds were created that consisted of approximately the same number of line segments as the backgrounds, but conveyed no meaning. Experiment 3 was a replication of Experiment 1 with the nonsense backgrounds in place of the no background controls. The results indicated that consistent backgrounds facilitated object identification while inconsistent backgrounds did not interfere.

These experiments, taken as a whole, seem to indicate that background information, as defined by Boyce et al. is responsible for the scene context effects, while cohort objects in the scene have no effect on the identification of the target object. However, other research exists that indicates that cohort objects in the scene may play an important role in facilitating object identification.

Friedman (1979) proposed that some objects in scenes may be responsible for facilitating further object identification. She argued that some objects are obligatory given the scene context (e.g. refrigerator in kitchen scene) while other objects are non-obligatory (e.g. plant in kitchen scene). Obligatory objects may provide access to memory for similar scenes which should facilitate further scene processing. Although she did not collect data to directly test the role of obligatory objects as a source of context effects, she has evidence for the distinction between obligatory and non-obligatory objects. Friedman recorded eye movements while subjects viewed scenes in preparation for a recognition memory test. She found that first fixation durations were shortest on obligatory objects and were longest on unexpected objects that did not fit in the scene at all. First fixations on non-obligatory objects were of intermediate duration.

A second source of support for objects as the locus of scene context effects comes from Henderson, Pollatsek and Rayner (1987). They proposed that objects might facilitate one another in scene context due to the semantic relationships between objects likely to occur in the same scene. They presented subjects with two objects that were to be fixated sequentially and recorded eye movement times and time to name the second object. They found that fixating a related object prior to the target object facilitated naming time for the target object. Thus, fixating a picture of a doctor before fixating a picture of a nurse facilitated the time to say "nurse". However, these objects were not located in scenes, but as isolated objects without context. Thus it is unclear how this relates to research on scenes.

Conclusions about the locus of scene context effects are hard to draw from the above review. Boyce et al. have evidence that global information gained from scene backgrounds is an important source of contextual information and that cohort objects provide little, whereas data from Henderson et al. indicate that object information may also serve as the locus of context effects. These findings may not be quite as contradictory as they first seem. The methodologies employed were quite different (this will be discussed in more detail later) and it is possible that both object and global background information play a role in scene context effects.

Mechanism of Facilitation

Two opposing theories exist about the mechanism of scene context effects. One theory, put forth by Biederman (1982), Friedman (1979), and others (Antes et al., 1981; Loftus, 1983) holds that a schema or frame for the particular scene is accessed early in scene viewing. The second theory holds that a much simpler priming mechanism can account for scene context effects (Henderson et al., 1987).

What is a schema?. The typical definition given in response to this question is that a schema is a hierarchically organized memory structure. This definition is accurate but fairly abstract. All the knowledge of the world that we have must be organized in our memory in some way. When presented with a concept such as "kitchen", we are able to access lots of information related to that concept fairly easily. We know that a kitchen is a room in the house generally used for cooking. Also, we know that utensils to accomplish this task (pots and pans, coffee makers, spoons, etc.) are all likely to occur in the kitchen. We know that stoves and refrigerators occur in kitchens and generally don't occur in other rooms of the house and so on. A schema is one way to think about how this information is organized. That is, information that is relevant to a particular concept is bundled together so that activating one concept makes it easier to access the concepts that are bundled with it.

One thing that makes schemas different from other organizations of memory (semantic networks, for example) has to do with what particular concepts are grouped together. A schema is organized around everyday events and scenes. Concepts that typically co-occur in an event or scene will be grouped together. Thus the basis of organization is "episodic" knowledge, not abstract, semantic knowledge.

A second characteristic of schemas is their hierarchical organization. This means that concepts that are episodically related but at different levels in category membership are organized such that more basic level concepts are grouped under higher order concepts. Thus, "bedroom" might be the top concept representation with objects that are likely to occur in a bedroom represented at a lower level and connected to the "bedroom" node.

A third characteristic of schemas is that for any particular concept there exists a default value. That is, when you read the word "spoon" probably the spoon that comes to mind is not a serving spoon or soup spoon, but the ordinary metal dinner spoon. That is, contained within the schema organization is also the information about which subordinate concepts are most frequent.

A scene schema is knowledge about a place that is organized in the manner outlined above. The superordinate concept usually is thought of as corresponding to the setting, such as "kitchen". Grouped under this concept are the representations for objects that are likely to occur in the scene and each object is also represented by its own schema. When one of the concepts in this structure becomes activated, this activation is spread to the other concepts that are connected to the activated node.

The critical difference between the schema organization and a semantic network, for the purposes of the current research, has to do with whether the relationships between the concepts are based upon episodic or semantic associations. Schemas predict that concepts that co-occur in the world are connected in memory. That is, the connections between representations in memory directly reflect our experiences in the world. Semantic networks, on the other hand, are organized around more generalized, abstracted knowledge and category structures.

To summarize, a schema is a possible organization of world knowledge in memory. The general principles that define a schema are a) concepts that co-occur in scenes and events in the real world are connected in memory, b) concepts are represented at many levels of specificity and arranged hierarchically, and c) there exists default values that usually represent the most common instantiation of the concept.

One final aspect of schema organization that is often included in the definition of schema is that schemas also contain information about the preferred spatial locations of objects within scenes. It has been proposed (Biederman et al., 1982) that somehow information about the spatial position of an object is contained in the connection between the object representation and the superordinate representation. There is no clear discussion in the literature as to how this information is represented in the schema. Since this aspect of schema organization is not very well understood and the experiments reported in this paper do not directly test the issue of spatial locations in schemas, I will not consider it further.

Schemas and Context Effects. If schemas are to be of value in accounting for the fact that scene context facilitates object identification, then one must explain how a particular schema gets activated and how this facilitates the process of identification. One possibility is that identification of global contextual information makes the first contact with the schema for that scene. Once the superordinate concept for the scene has been activated, the rest of the concepts grouped together in the schema for that scene become partially activated. The data from Boyce et al. (1989) and from Loftus and Mackworth (1978) would argue that global scene context activates the appropriate schema. A different route to schema activation has been proposed by Friedman (1979). She suggests that the identification of an obligatory object in the scene accesses the correct schema, which then speeds up the identification of the remaining objects.

Both of these routes to schema activation could be correct. That is, it may not matter what aspect of the scene is identified first, whether it is the scene background or a particular key object. The important finding is that something about having an object in a scene seems to facilitate object identification and this facilitation may be accomplished by way of activating the appropriate scene schema.

Object to Object Priming

A very different mechanism of facilitation has been proposed by Henderson et al. (1987). They argued that facilitation occurs between related objects by a process of automatic, spreading activation, whereby objects in a scene that are related semantically or associatively facilitate one another. The difference between this view and that put forth by the schema theorists is the prediction of what should facilitate what. That is, if objects are connected to one another in a pre-existing semantic network as proposed by Henderson et al. and facilitation is the result of spreading activation through this network, then it is unclear how higher-order conceptual information can produce facilitation in object identification. In this scheme, context effects can only occur from activating a close neighbor of a particular object in the network. The advantage of this theory is that it makes much clearer predictions concerning which objects should prime each other. Objects that are semantically related, such as "coat" and "hat" should facilitate one another. Also, objects that are strong associates of one another should result in facilitation, such as "doctor" and "nurse".

However, the validity of object to object priming as a theoretical explanation of scene context effects rests on its ability to explain facilitation effects found with normal scenes. The assumption is that objects that are strong associates of one another or are semantically related are commonly found in the same scene. While it is true that coats and hats can appear in a scene together, I would argue that there is a different level of relationship between objects in most scenes. Many objects in scenes are _episodically related_, meaning that they can co-occur given the setting information but a priori would not be strongly related to one another. For example, both a teddy bear and a suitcase could appear in a scene sitting on a bed. These objects can co-occur in the bedroom scene, yet they are not semantically or associatively related.

To summarize, work on scene context effects has focused on the issue of what in the scene is responsible for the facilitation of object identification and how this facilitation occurs. The research conducted by Boyce et al. (1989), Loftus and Mackworth (1978), Friedman (1979) and Biederman et al. (1982) argues that a schema for the scene is activated early in scene viewing and this facilitates further scene processing. There is little agreement or specificity concerning what aspects of the scene are responsible for this facilitation. However, much of the evidence indicates that global information about the setting of the scene may be at least partly responsible.

Methodological Problems with the Previous Research

Most of the research discussed in the previous section has been conducted with either brief presentation using accuracy as the primary measure or with eye movement monitoring techniques. The rationale behind the brief presentation is to simulate the first fixation on a scene. The durations of the display in this research range from 100 to 200 ms. This is approximately the length of first fixations on scenes when not presented tachistoscopically (Loftus, 1983).

There are two possible problems with brief presentation research. First, one could argue that this methodology may overestimate the effect of context during normal scene viewing. When a scene is presented for 100-200 ms, subjects do not have time to identify all of the scene elements. In the cases where the subject has not identified the target object during the scene exposure, the presence/identity of the target object may be inferred from what the subject knows is likely to occur, given his world knowledge and some partial scene information.

Second, it is possible that brief presentations are not appropriate simulations of one fixation on a scene. It could be that subjects employ attentional strategies with t-scope displays that they would not use in normal scene viewing. When the scene is presented for 100 ms and the subjects have no idea where the critical region of the picture is prior to the trial, they may try to diffuse their attention over the entire visual field. This strategy could be beneficial, since some information should be obtained from most regions of the scene which may outweigh the cost of not processing any one location of the scene very well. However, this same strategy may not be so beneficial when the viewing time is longer and there is not such a pressing need to process all areas of the scene at once.

The other paradigm used in this research has been eye movement monitoring. Typically this is done while subjects view scenes for longer durations (approximately 2-10 seconds). Commonly, subjects are instructed that there will be a recognition memory test for the pictures. The dependent variables of interest are fixation duration and fixation location. It is thought that subjects will fixate informative regions of the scene early and that fixation duration will reflect object identification time. There are also some problems with this technique. First, studying a picture for a later recognition memory test may be very different than other types of viewing. It is possible that subjects pay more attention to small details of a scene than they would if a different task had been employed. The second problem with this research is the validity of the dependent measures. Although fixation duration has been demonstrated to be a correlate of word identification time in reading (Blanchard, 1985; Rayner and Pollatsek, 1987) the situation could be different for pictures. It is possible that duration of a fixation on an object could primarily reflect higher order factors such as subjects' interest in the particular object rather than time to identify the object.

The Current Methodology

The best way to investigate context effects on object identification would be to allow subjects enough time to view the scenes "normally" and use some measure of object identification time besides fixation duration. The procedure used in the following experiments approximates these requirements. The basic idea is that subjects fixate two locations in a scene, with the first fixation in the center of the display and the second on a target object. Object identification time is assessed by measuring the time to name the target object. The procedure is similar to that used by Henderson et al. (1987), and Pollatsek, Rayner and Collins, (1984) except that the objects were embedded in normal scene context.

However, since there are many objects in a given scene, a cue must be given to indicate which object was the target object. Therefore, during the first fixation on the scene (in the center) one of the objects is moved (approximately 1/2 degree) and quickly returned to its original position. This movement is easily perceived by the subjects and serves as a location cue for the target object. The subject's task is to move his or her eyes to the target object and name it as quickly as possible.

This procedure has some important advantages over t-scope presentation. Since the scene was present for as much time as the subject needs to complete the task, there should be no need for the subject to invoke atypical attentional strategies. Also, since the subject always has time to fixate the target object there should be no need for the subject to rely on context more in this task than in normal scene viewing. That is, subjects will not have to infer the identity of the target object post-perceptually, since a trial does not end until the subject has named the object.

The dependent measures -- naming time, fixation durations and fixation locations -- together should be a better reflection of object identification processes than fixation duration alone, as has been previously used. Time to name an object seems a reasonable correlate of actual identification time given current theories of object recognition (Kroll and Potter, 1984; Theios and Amrhein, 1988). It has been proposed that an object must be identified conceptually in order to access the stored name, whereas with words it is possible to produce the name without necessarily contacting the meaning.

The research reported in this paper was conducted to further our understanding of the effect of scene context on object identification. The purpose of the first experiment was to determine whether scene backgrounds facilitated object identification with a different task. That is, it was of interest whether given good perceptual information about the scene and the target object (i.e. longer viewing duration, opportunity to fixate the target object) context would still affect object identification. This is important since the bulk of experimental evidence for scene context effects is from the t-scope paradigm.

The second and third experiments assessed the time course of contextual information acquisition. The experiments tested whether context was processed a) only on the first fixation on the scene, b) mostly on the second fixation on the scene or c) during both fixations on the scene. Through an understanding of the time course of contextual processing we can better understand the mechanism by which context facilitates the object identification process.

CHAPTER 2

EXPERIMENT 1

Introduction

The purpose of the first experiment was to determine whether scene context affects object identification using a more controlled methodology than has been previously employed. As mentioned in the introduction, there are a number of reasons to question the body of literature on context effects because of the methodology used. Evidence from tachistoscopic displays may overestimate context effects and previous eye movement research with objects in context may have used a contaminated dependent measure. This experiment attempts to determine whether context affects object identification when the methodological problems are eliminated.

Method

Subjects

Eight graduate and undergraduate students from the University of Massachusetts participated in the experiment. The subjects were paid $15 for their participation and the experiment took 3 one-hour sessions.

Stimuli

There were 64 scenes used in the experiment, which were constructed from an original set of 16 scenes. These 16 scenes were line drawings of rooms of a house (e.g. bedroom scene), common public places (e.g. diner/lunch counter), and common outdoor scenes (e.g. street scene). (See Appendix for a list and description of scenes). Each of the original 16 scenes consisted of a background and five objects that belonged in the background. These 16 scenes were modified to create the Consistent Background scenes, the Inconsistent Background scenes and the Nonsense Background scenes.

The original 16 Consistent scenes were organized in 8 scene pairs such that objects in both scenes in the pair were roughly equivalent in real-world size. The 16 Inconsistent Background scenes were created by switching objects in one scene with objects in the paired scene. The placement of objects in the Inconsistent background was subject to the requirement that the objects not appear unsupported (which required rearranging the objects in many cases). In no case was the actual size of objects altered. However, since the objects were in different locations in the Consistent background and Inconsistent background conditions, Nonsense Backgrounds were used to control for the effects of this object relocation.

Nonsense backgrounds were created by distorting the original backgrounds in order to delete the semantic information contained in the background. Distortions were created with the following criteria in mind: a) nonsense backgrounds should not look like the original backgrounds; b) subjects should not be able to name these backgrounds; c) roughly the same number of lines and angles should be employed in the nonsense background as in the original; d) and nonsense backgrounds should preserve a three dimensional quality. The nonsense backgrounds resulting from this set of criteria did not appear as a random set of line segments, but as a "coherent" background that lacked theme information. The three-dimensionality of the nonsense backgrounds was preserved in order to provide planes on which objects could be supported. The objects were situated in the nonsense backgrounds in exactly the same places as they were in their matched meaningful background. Thus, the effects of distance from the fovea could be assessed independently of semantic content of the background. Each original background had a nonsense background version.

The end result was 16 background consistent scenes, 16 background inconsistent scenes and two sets of nonsense background scenes of 16 scenes each. The two sets of nonsense scenes were the result of placing objects in the nonsense backgrounds to match the spatial locations of consistent scenes and inconsistent scenes. The scenes subtended approximately 15 degrees visual angle and the objects subtended approximately 2 degrees visual angle. These stimuli were exactly the same as used by Boyce et al. (1989), Experiment 3.

Design

Each subject named each target object in all four background conditions (consistent, inconsistent and two nonsense controls). Each scene contained five objects, four of which served as target objects to be named. In each scene, the object that was most difficult to name in a prior pilot study was not included as a target object. Thus, in the experiment, each of 64 target objects appeared in four conditions, resulting in 256 trials. Trials were presented in a random order.

Apparatus

The scenes were displayed on a Megatek Whizzard vector plotting CRT with a P-31 phosphor. This was interfaced with a VAX 11-730 computer. Short vectors (under 1/4 degree visual angle) were plotted in 1.2 microseconds per vector and long vectors were plotted in 2.0 microseconds. Total plotting time for an average scene was under 4 ms. Eye movements were monitored with an SRI Generation V Dual Purkinje eye tracker interfaced with the computer. Naming times were measured by a standard switch closure voice key (Gerbrands). Subjects' naming times, eye positions, and fixation durations were collected and stored by the computer.

Procedure

Subjects were initially shown the set of 64 target objects, one at a time, in isolation, in order to establish names for the objects. A pilot study indicated that most subjects used the same names for the objects. However, subjects were allowed to use any name they felt comfortable with, provided they were consistent across trials. After the naming practice, subjects placed their heads on a chin rest in front of the CRT and their eye movements were monitored. When the experimenter determined that the eye tracker was accurately monitoring the subjects' eye position, the calibration sequence began. Subjects sequentially fixated crosses located on the left, right, top and bottom of the screen. After a successful calibration had been achieved, subjects were given instructions about their task. Ten practice trials preceded the experimental trials in order to demonstrate the "wiggle" to the subject. Different scenes and objects were used in the practice trials than in the experimental trials.

A trial began with a fixation cross in the center of the screen. When the subject was seen to be fixating the fixation cross, the experimenter began the trial. The scene appeared on the screen. After 75 ms. one of the objects moved 1/2 degree visual angle and returned to its original position 75 ms later. Subjects moved their eyes to the "wiggled" object and named it as rapidly as possible and the activation of the voice key terminated the display. The experimenter recorded the response as correct or incorrect. The experiment consisted of 256 trials separated into 8 blocks of 32 trials each. A trial block took approximately 15 minutes to complete so the entire experiment spanned three one hour sessions conducted on consecutive days.

Subjects were given a 5 minute rest between blocks run on the same day.

Eye Movement Reduction Criteria

Typically, eye movement researchers have used the term "fixation" to refer to the period of time the eye is in a stable position and the term "saccade" to refer to the period of time the eye is in motion. Use of these terms implies that there is an unambigious criterion for the determination of fixations and saccades in the eye movement record. However, this is not the case. Some criterion must be established to draw the line between the end of a fixation and the beginning of a saccade. A variety of different criteria have been used in the eye movement literature. In the current experiment, the eye was considered in motion (saccade) if the position of the eye in a given millisecond was more than 1/2 degree visual angle different than the average of the preceding 5 ms. Thus, small movements, under 1/2 degree, were not considered true saccades and were included as part of the fixation duration. Large movements, greater than 1/2 degree were considered saccades.

Results and Discussion

The primary dependent measure was the time to name the target object. Eye movement data were collected primarily to verify that subjects were performing the task as instructed: that they began a trial fixating the center cross and then moved their eyes directly to the wiggled object. The eye movement data allows one to exclude trials on which the subject did not move their eyes directly from the center of the scene to the target object. The eye movement record also allows one to measure naming time in a number of different ways. Instead of measuring naming latency from the beginning of the trial until the voice key was activated, naming time may also be measured from the time that the object was initially fixated until the activation of the voice key. The data from the current experiment was analyzed two ways: by taking the eye movement data into account and also by ignoring the eye movement data. Both will be reported here.

Total Naming Time Analysis

The mean object naming latency of all correctly named objects can be seen in the Table 1 below. Naming errors occurred on less than 1% of the trials. The primary goal of Experiment 1 was to assess the effect of background information on object naming. To this end, one might want to compare the consistent background condition directly to the inconsistent background condition. However, since the objects were in different locations in these two conditions, the appropriate measure of the background effect is the interaction between the meaningfulness of the background and the consistency of the background. That is, since the nonsense background conditions were matched to the experimental conditions for placement of the objects, a more appropriate comparisons are between the meaningful background conditions and their matched nonsense background conditions.

Table 1

Mean Object Naming Time Measured from the Beginning of the Trial in Experiment 1 (in milliseconds)

	Consistent Background	Inconsistent Background
Meaningful Background	1082	1146
Nonsense Background	1128	1125
Differences	-46	+21

Note - Includes all data except for naming errors. No data have been excluded due to the eye movement record.

The overall naming latency in the experiment was 1120 ms. As can be seen in Table 1, objects were named 46 ms faster in the Consistent Background condition than in the matching Nonsense Background control condition and objects in Inconsistent scenes were named 21 ms slower than objects located in the matching Nonsense Background condition. The interaction between Background Consistency and Background Meaningfulness was significant, $F(1,7) = 7.88$, $p < .026$, $MSE = 1152$. There also was a significant main effect of Consistency, $F(1,7) = 31.64$, $p < .001$, $MSE = 234$.

In order to evaluate whether the consistent backgrounds facilitated object identification and whether inconsistent backgrounds interfered with identification of the target, simple effects t-tests were computed for the differences between the experimental conditions and their matching controls. Both the -46 ms difference between the Consistent Background condition and its Nonsense Background control and the 21 ms difference between the Inconsistent Background condition and its Nonsense Background control were significant, $t(7) = -2.14$, $p < .05$, ($s = 61$), and $t(7) = 2.24$, $p < .05$, ($s = 27$), respectively.

To summarize, when considering overall naming time from the beginning of the trial until the object was named, there was a significant effect of background consistency on object naming relative to the nonsense background controls. Furthermore, from this analysis it appears as though consistent backgrounds facilitated object identification and that inconsistent backgrounds interfered with the object identification process.

Corrected Naming Time Analysis

Criteria for Excluding Trials. The data included in the following analyses are essentially the same as that presented above, with the exception that some data were excluded on the basis of the eye movements. In order for a trial to be included in these analyses, the subject must have begun the trial fixating within 1 degree visual angle of the center fixation cross. In addition, the eye must have remained in this center region of the screen until the wiggle occurred and then the next fixation longer than 50 ms must have been within 1 degree of visual angle of the target object. The 1 degree cut-off was chosen because it was stringent enough to determine unambigiously that the subject was fixating the target object (and the center of the scene at the beginning of the trial) and not another object in the scene, yet loose enough that trials were not excluded due to minor errors in calibration. The requirement that the eye move directly from the center to the target object region with only fixations of less than 50 ms in between was chosen because it allowed no meaningful fixations on other objects to intervene between the first fixation and the target object fixation; however, it does not throw out trials in which the subject blinked or there was a short temporary track loss.

Corrected Overall Naming Time. The corrected overall naming time data after the above cut-offs have been applied can be seen in Table 2. The mean naming latency averaged across conditions was 1109 ms.

Table 2

Mean Overall Corrected Naming Time in Experiment 1 (in milliseconds)

	Consistent Background	Inconsistent Background
Meaningful Background	1073	1129
Nonsense Background	1117	1118

Note - Eliminating trials where the eye did not begin in the center of the scene, remain in the center of the scene until after the wiggle, or fixated some other aspect of the scene for longer than 50 ms. before fixating the target object.

The difference between the Consistent Background condition and its Nonsense Control was -44 ms and the difference between the Inconsistent Background condition and its matching Nonsense Background control was 11 ms. This interaction was significant F(1,7) = 5.60, $p < .048$, MSE = 1074. There was also a significant main effect of Background Consistency, F(1,7) = 11.3, $p < .011$, MSE = 562. Simple effects t-tests were conducted to determine whether the naming time in the Consistent Background condition was significantly faster than in the Nonsense Background control and whether the naming in the Inconsistent Background condition was significantly slower than in its control condition. The difference between the Consistent Background and the Nonsense Consistent condition was -44 ms, t(7) = -2.41, $p < .025$, (s = 52) and the difference between the

Inconsistent Background condition and its Nonsense Background control was 11 ms, $t(7) = 1.28$, $p > .10$.

Thus, the overall pattern of data from this analysis looks similar to the uncorrected naming analysis with the exception that in this analysis there is stronger evidence for facilitation and weaker evidence for interference in object naming caused by the inconsistent backgrounds.

Naming Time from Fixation of Target Object. Another way to conceive of object naming latency would be to use the time between when the object was actually fixated and when it was named. The advantage of measuring naming time in this manner is that it may more truly reflect the processing required to identify the target object. Table 3 displays the mean object naming latency, measuring from when the target was fixated to the end of the trial (triggering of the voice key).

Table 3

Mean Naming Time Measured from when Target Object was Fixated Until End of Trial in Experiment 1 (in milliseconds)

	Consistent Background	Inconsistent Background
Meaningful Background	751	814
Nonsense Background	798	806
Differences	-47	+8

Mean overall naming latency in Table 3 is 792 ms. The critical interaction between Background Type (meaningful or nonsense) and Background Consistency was significant, F(1,7) = 6.17, p < .041, MSE = 984, as was the main effect of consistency, F(1,7) = 19.6, p < .004, MSE = 517. These effects are consistent across items as well. In an items analysis the interaction was significant, F(1,63) = 6.44, p < .013, as was the main effect of consistency, F(1,63) = 6.35, p < .014.

The difference between the Consistent Background condition and its matching Nonsense Background control was -47 ms, which was significant, t(7) = -2.56, p < .025,(s = 52), and the difference between the Inconsistent Background condition and the Nonsense control, 8 ms, was not significant, t < 1.

To summarize, again there is evidence that consistent backgrounds facilitate the object identification process, while there is no clear evidence for interference by the inconsistent backgrounds. Taken as a whole, these different analyses on the naming time data definitively point toward the conclusion that there is an effect of background consistency on the latency to name the target object. Furthermore, in all analyses there is evidence for facilitation from the consistent backgrounds and there is less evidence that inconsistent backgrounds interfere.

The results of the analyses on the naming time data, taken as a whole, indicate that 1) context does affect the process of object identification and 2) consistent context facilitates the object identification process while inconsistent context interferes little. Thus, the same basic effects that were obtained in Boyce et al. (1989), Experiment 3 have been replicated with a different methodology.

In the introduction, it was suggested that context effects may not be evident given a longer viewing time. That is, if the subject has ample time to view the object that the identification process would work so efficiently that context would not play a role. The evidence does not support this view. In this experiment viewing times on the scenes were approximately 1 second and errors in naming were less than 1%. Obviously this is enough time for the background context and the target object to be identified.

Eye Movement Latency

The final dependent measure from Experiment 1 is the latency of the first eye movement. This measure is the time it took subjects to begin an eye movement from the center of the scene to the wiggled object. It was thought that differences in this measure might reflect processing of background information on the first fixation on the scene.

Table 4

Mean Eye Movement Latencies in Experment 1 (in milliseconds)

	Consistent Background	Inconsistent Background	Average
Meaningful Background	301	293	297
Nonsense Background	296	289	293
Average	299	291	

In Table 4 one can see the mean eye movement latencies for each of the conditions in Experiment 1. Overall the mean eye movement latency was 295 ms. The only significant effect present in these data was a main effect of Consistency $F(1,7) = 7.96$, $MSE = 57$, $p < .025$.

This is a somewhat puzzling result. One might expect, with this measure, to find an effect of background meaningfulness. It would seem logical that subjects would be either faster or slower to move their eyes from center given a nonsense background. However, there was no significant effect of background meaningfulness but there was an effect of background consistency. Could it really be the case that subjects can identify whether the objects are consistent or inconsistent with the background completely on the first fixation on the scene? It seems likely that there is a slightly more mundane explanation for the main effect of consistency found in the eye movement latency data. Recall that the objects are in somewhat different spatial locations in the consistent background and inconsistent background scenes. Although overall the objects are equidistant from the center of the scene, it seems likely that something about the placement of the objects in the scene frame is responsible for this effect. That is, the location of the objects in the inconsistent backgrounds are somewhat easier to respond to when the target object wiggles. Although the effect is significant the absolute value of the difference is small (8 ms).

Since this effect is significant, however, it seems as though the best measure of the object identification time is the naming time from when the target object is fixated. Both of the other two overall measures include the time to move off the center of the object which is possibly confounded with the position of the object.

Subsidiary Analyses

Practice Effects. In order to determine whether the above reported context effects are attenuated over the course of the experiment the data from this experiment was divided in half and a three way Analysis of Variance was performed with first-half/second-half entered as a factor in the design. Overall, subjects were 86 ms faster in the second half of the experiment than in the first half. This difference was significant, $F(1,7) = 43.32$, $p < .001$. However, no other factor interacted with practice. That is, although the overall naming time was faster in the second half of the experiment, the difference between the Consistent Background condition and the Inconsistent Background condition stayed roughly the same (Consistent - Inconsistent in the first half = -51 ms. Consistent - Inconsistent in the second half = -71 ms.). However, since the order of the trials was completely randomized across the experiment, this analysis is confounded with items.

Distance Effects. Another subsidiary analysis of interest is whether the context effect stays the same whether or not the target object is close to the center of the scene or far away from the center of the scene. The items were divided into two groups: a near group, which was approximately 3 degrees visual angle from the center, and a far group, which were approximately 7.5 degrees from the center. The difference between the Consistent Background condition and the Inconsistent Background condition in the near was only -8 ms, while the same comparision for the far items was -212 ms. Furthermore, the interaction between background consistency and background type (either meaningful or nonsense) was not significant for the near items, $F(1,7) = .06$, $p < .79$, while this interaction was significant for the far items, $F(1,7) = 9.35$, $p < .018$. From this it seems reasonable to conclude that distance of the target object from the center does play a role in the degree to which context will affect identification. However, this analysis is somewhat confounded with object since each object does not occur in both near and far locations. Also, some of the cell means in the control conditions appear odd, probably due to the small sample size.

Correlation between Consistency Effect and Independent Ratings. The purpose of this analysis was to determine how well the obtained consistency effect covaried with how predictable the object was from context. A separate group of subjects was given verbal descriptions of the scenes and then the names of the objects that appeared in the scenes. These subjects were instructed to rate the object on a scale from 1 to 6, with 1 meaning very unlikely to occur in the scene and 6 meaning very likely to occur in the scene. These ratings for each item were then correlated with the consistency effect for that item (Background Consistent - Background Inconsistent). The correlation between these ratings and the consistency effect was relatively low, $r = .20$, but marginally significant, $p < .10$.

It had been previously thought that context effects would be evident only when the objects are highly predictable from the context. This result indicates that context affects object identification even when the object is only plausible in the context, not highly predictable from the context.

In summary, consistent scene context facilitates the object identification process. Furthermore, there is little evidence that the inconsistent scene context interferes with object identification. Objects located close to the original fixation are less affected by context. Presumably this is because when the target object is so near, it is identified in parallel with the background. Identification of the target in this case may be completed before the identification of the background information. Finally, it has been found that although there is a moderate correlation between an object's predictability and its context effect, this by no means accounts for the context effect completely. That is, it appears as though objects only need to be in a plausible context for facilitation to occur, not a highly predictive context. This result mirrors the results of Boyce et al. (1989).

Conclusion

These data considered together with previous research with these scenes indicate that the background information is certainly processed early enough to facilitate object identification. In fact, results from the t-scope experiments suggest that the meaning of the background can be extracted in a 150 ms view of the scene. The current study, however, is not diagnostic about the time course of background information acquisition. It could be that the background information was obtained primarily on the first fixation on the center of the screen and then subsequently ignored. Or it could be the case that information concerning the background is continuously processed throughout the duration of the scene viewing. The time course of background information utilization is the question the second experiment attempted to address.

CHAPTER 3

EXPERIMENT 2

Introduction

So far it has been demonstrated that scene context does facilitate object identification even with relatively long viewing times. The evidence from Experiment 1 and the evidence from previous scene research with shorter viewing durations suggests that scene context information is extracted quite early in scene viewing. In fact, the evidence so far may lead one to suggest that the first stage of scene processing is the extraction of global scene meaning and processing shifts to the objects contained within the scene only when that stage is complete. The purpose of Experiment 2 was to test claims about the time course of contextual information utilization.

Loftus (Loftus and Mackworth, 1978; Loftus, 1983) has suggested that the scene context is identified on the first fixation, and this guides subsequent scene processing. Specifically, he claims that the activation of the correct schema for the scene allows one to determine the likelihood of occurence of the objects located in the scene. Then the eyes are guided to objects that have a low probability of occurring in the scene.

However, evidence for this view that scene context is acquired during the first fixation comes, not from his eye movement research, but from conclusions drawn about t-scope research. He argues that if context effects can be demonstrated when the scene has been viewed for only 150 ms then contextual information must be extracted during the first 150 ms of scene viewing. While it is true that information about the context of the scene can be accessed in 150 ms (as evidenced by the t-scope research) this does not necessarily argue that given longer viewing time all of the needed contextual information is extracted in the first 150 ms.

If Loftus is correct and all the relevant contextual information is extracted during the first fixation on the scene, then the presence of context information on subsequent fixations on the scene should be relatively unimportant. Experiment 2 was a test of this claim. In this experiment the background information present on the first fixation of the scene changed contingent on the position of the subject's eye. That is, the type of context varied from the first fixation to the second fixation on the scene in an attempt to identify when during scene viewing context information is utilized.

Method

Subjects

Nine University of Massachusetts graduate and undergraduate students participated in the experiment for payment. Subjects were paid $25 dollars upon the completion of the experiment for a total of 4 hours of participation. None of these subjects had participated in Experiment 1.

Stimuli

The scenes were exactly the same as those used in Experiment 1. However, not all of the target objects from the first experiment were tested in the current experiment. In order to reduce the number of trials in this experiment, three target objects were chosen for each scene instead of four as in the previous experiment. (Thus, two objects were not cued from each scene.) The decision about which additional object to exclude was based on a number of criteria: a) if subjects had difficulty naming it in Experiment 1; b) because the voice key was unresponsive to the name in the first experiment; and c) if the object was extremely close to the center. The latter criterion was invoked to ensure that the saccade time would be sufficiently long to allow for completion of the display change. These objects were not eliminated from the scenes; they just were not cued at any time during the experiment.

Design

The background information was either consistent, inconsistent or nonsense on the first fixation in the center of the scene and was either consistent, inconsistent, or nonsense on second fixation on the scene. Because the objects were located in different places in the scenes in the consistent and inconsistent backgrounds, it was not possible to directly change a consistent background into an inconsistent background (or vice versa) in a given trial. Instead, all of the comparisons had to be made against the relevant nonsense background control conditions. This resulted in the eight conditions shown in Table 5.

Table 5

<u>Conditions of Experiment 2</u>

	First Background Type	Second Background Type	Label
a)	Consistent	Consistent	BC/BC
b)	Consistent	Nonsense	BC/NBC
c)	Nonsense	Consistent	NBC/BC
d)	Nonsense	Nonsense	NBC/NBC
e)	Inconsistent	Inconsistent	BI/BI
f)	Inconsistent	Nonsense	BI/NBI
g)	Nonsense	Inconsistent	NBI/BI
h)	Nonsense	Nonsense	NBI/NBI

Conditions a,d,e, and h, where the same scene was present on the first and second fixations on the scene, were essentially replications of the conditions of Experiment 1. However, for purposes of control, display changes were enacted with these four conditions as well. There might have been some visual disruption associated with changing the background, even though the change occured during the saccades. Therefore, display changes were carried out on 100% of the trials. In the above four conditions where the content of the background remained the same across the two fixations, the second version of the background was reduced in size by 10%. It was hoped that this minor size change would cause similar surface disruptions as in the four other conditions.

Each of the 48 target objects appeared in all eight conditions, resulting in 384 trials per subject. The presentation of the trials was randomized across the entire experiment. The set of 384 randomized trials were divided into 8 trial blocks of 48 trials a piece. Subjects took four one-hour sessions to complete the experiment.

Apparatus

The equipment used in Experiment 2 was exactly the same as in Experiment 1.

Procedure

The procedure of this experiment was similar to that of Experiment 1. The only difference was the display change that occurred during each trial in this experiment. The display was changed when the subject began an eye movement off the center of the scene. When subject's eyes crossed an imaginary boundary 1 degree around the fixation cross the currently displayed background was removed from the screen and a second background was plotted on the CRT. This change of backgrounds was accomplished within 10 ms, which was faster than the duration of a typical saccade. Due to the suppression of visual information processing during saccades, the subject did not see the display change occurring. When the subject's eye landed on the target object, the object was now in a different context than when the trial began. The voice key was triggered when the subject named the target object and this terminated the entire display.

Results and Discussion

The purpose of Experiment 2 was to investigate the time course of context processing. Varying whether or not a consistent, inconsistent or nonsense context was present on the first, second or both fixations seemed a reasonable way to investigate this question. The questions of interest were whether a) all relevant contextual information is processed early in scene viewing or, b) whether contextual processing continues in parallel with the identification of the target object. As with the previous experiment, comparisons cannot be made between Consistent Background conditions and Inconsistent Background conditions because the objects were in somewhat different locations in the two types of backgrounds. All comparisons must be made against the appropriate nonsense background condition.

As with Experiment 1, there were a number of ways to measure naming time and also many ways to decide which trials to exclude. In Experiment 1 there were few differences between the various measures. In this experiment, only one measure of naming time will be reported, although the other measures were computed. Again no differences between the measures were found, so the naming times reported here were measured from when the target object was fixated until the target was named. A trial was included in this analysis if the subject a) began the trial fixating within 1 degree of the center fixation cross and b) eventually fixated the target object.

Table 6

Mean Naming Time in Experiment 2 (in milliseconds)

		Background on 2nd Fixation	
		Consistent Background (BC)	Nonsense Background (NBC)
Background on 1st Fixation	Consistent Background (BC)	663	682
	Nonsense Background (NBC)	684	673

		Background on 2nd Fixation	
		Inconsistent Background (BI)	Nonsense Background (NBI)
Background on 1st Fixation	Inconsistent Background (BI)	724	702
	Nonsense Background (NBI)	732	717

Note - Naming time measured from when target object was fixated until target object was named.

The means for the eight conditions can be seen in Table 6. The overall naming time averaged across conditions in Experiment 2 was 697 ms. Subjects made errors in naming the target object less than 1% of the time.

Overall, a context effect can be seen in the data. The conditions where an inconsistent background was present (BI/BI, BI/NBI, NBI/BI) resulted in longer naming times than the consistent background conditions (BC/BC, BC/NBC, NBC/BC), 719 ms vs. 676 ms, respectively. The main effect of background consistency was significant, $F(1,8) = 47.7$, $p < .0003$. Furthermore, the context effect in this experiment (BC/BC - BI/BI, a 61 ms difference) was about the same in magnitude as the context effect in Experiment 1 (63 ms). A major problem with the data, however, is that a difference was found between the NBC/NBC condition and the NBI/NBI condition (a 44 ms advantage for NBC) and this difference was significant $t(8) = 3.14$, $p < .05$, $(s = 42)$. In Experiment 1 no difference was found between these two conditions. Theoretically, these should be equal because there should have been no consistent or inconsistent interpretation of the nonsense backgrounds with respect to the target object. There were no other significant effects.

The consistency effect found with the Nonsense Background conditions makes any of the other comparisons hard to interpret. For example, it could be the case that the BC/BC condition and the BC/NBC condition were not significantly different from one another because there was no additional advantage when the consistent background was present on the second fixation. This would argue that background information is acquired only during the first fixation on the scene. However, these two conditions may be the same because for some reason the nonsense backgrounds conveyed meaning about the scene context and so performance in the BC/NBC condition was the same as in the BC/BC condition.

It may be that the reason context effects were found with the nonsense controls has to do with the nature of the design of this experiment. In all cases where the background information changed in content (BC/NBC, NBC/BC. BI/NBI, NBI/BI) the change was between a meaningful background and the nonsense background that was a distortion of the meaningful background. It could be that over the course of the experiment subjects began to learn this pairing. The result of this would be that the nonsense backgrounds would no longer be meaningless, but would be interpreted as a version of the original background. For example, perhaps the nonsense version of the street background was no longer a meaningless context but instead represented the "messed up version of the street scene" to the subjects. A few of the subjects voluntarily expressed this strategy at the end of the experiment.

If this learning explanation were the cause of the difference between the nonsense conditions, then doing an analysis of practice effects should reveal different patterns of data at the beginning and end of the experiment.

That is, if subjects were learning the pairing between the nonsense scenes and the meaningful scenes then perhaps there should be no difference between the nonsense controls in the beginning of the experiment. Unfortunately, since the order of trials was completely randomized across the experiment any practice analysis will be confounded by item effects.

One way to analyze the effect of practice on this difference between the control conditions was to simply divide the experiment into two halves. The first set of four trial blocks makes up the first half and the second set of four trial blocks makes up the second half. The difference between the NBC/NBC and the NBI/NBI conditions in the first block was 54 ms and in the second block was 60 ms. This does not support my contention that subjects learned the pairings over the course of the experiment. However, one must keep in mind that since trials were presented in a completely randomized order, this analysis is confounded with items.

Perhaps a better way to assess practice effects would be to analyze only the data from the first experience with each target object for each subject. In this way, hopefully some of the item effects could be controlled for. However, this results in a very small number of trials. Therefore, means were computed counting only the first time the subject saw a particular target object in the consistent backgrounds and the first time they saw the target in the inconsistent backgrounds. Since order of items was not conterbalanced across condition it resulted in a somewhat unequal N per condition. The results of this analysis were not much more successful than the previous practice analysis. While a difference between the nonsense backgrounds still exists (NBC/NBC was 35 ms faster than NBI/NBI), the difference is smaller than the overall difference.

To summarize, there is not very strong evidence for the proposal that subjects learned the pairings between the meaningful backgrounds and nonsense backgrounds. However, the above practice analyses were all confounded in one way or another. The fact remains that context effects were found with the nonsense backgrounds (in contrast to Experiment 1) and this made other differences (or lack of) in the experiment impossible to interpret.

One interesting facet of this experiment was the number of times subjects reported an awareness of a scene change. Scene backgrounds changed on 100% of the trials. On 50% of the trials this change was minor (only a size change) and on the remaining 50% of the trials the change was major (change in the meaning of the scene). When asked at the end of the experiment whether they noticed these changes, subjects drastically underestimated the percentage of times a change occurred. Most subjects thought the backgrounds had changed on 10-25% of the trials indicating that they were unaware of the display change on a large percentage of the trials.

In summary, because of the problems with the control conditions in this experiment, little can be concluded from these data. While the context effect found in Experiment 1 was replicated, we have learned little about the time course of contextual information processing. Experiment 3 was conducted to remedy this.

CHAPTER 4

EXPERIMENT 3

Introduction

The purpose of Experiment 3 was to evaluate the questions about time course of contextual processing set out in Experiment 2. Experiment 2 failed to illuminate the time course issue because the Nonsense Background Control conditions also showed a context effect. For this reason, better control conditions were needed.

A decision was made to use a grid background instead of the nonsense backgrounds to serve as controls for the differences in object location between the consistent and inconsistent scenes. The grid was a neutral background which conveyed no semantic or depth information. The major advantage of the grid was that no meaning could possibly be ascribed to it. Therefore, any observed difference between the "consistent" and "inconsistent" grid conditions would be due to the somewhat different spatial locations of the objects in these two conditions.

The major disadvantage of using the grid instead of the nonsense backgrounds is that the grid probably does not represent a neutral baseline upon which to measure facilitation versus interference. The grid differs from the meaningful backgrounds in many visual characteristics. For example, it contains no depth information and appears as a visually regular field, instead of a meaningless background.

For these reasons it was not possible with this experiment to determine whether consistent backgrounds facilitate object identification or whether inconsistent backgrounds interfere with object identification. However, information about the utilization of background information over time should be available from this experiment.

Method

Subjects

Eight undergraduate students participated in this experiment. Many of the students attended other universities and were spending the summer in Amherst. Upon completion of the experiment subjects were paid $30 for approximately 4 hours of participation.

Stimuli

The consistent and inconsistent backgrounds and the objects were exactly the same as were used in Experiment 2. However, the nonsense backgrounds were not used in this experiment and were replaced by Grid Backgrounds. The grid was constructed from twelve regularly spaced horizontal lines and twelve vertical lines. This grid background resembled a piece of graph paper. Objects were superimposed on the grid background and the portions of the grid that were located "behind" the objects were deleted. The resulting scene looked like randomly placed objects in a graph paper background.

Design

The design of Experiment 3 was virtually identical to Experiment 2. Subjects named each target object in each of the eight conditions shown in Table 7.

Table 7

Conditions of Experiment 3

	Background on Fixation 1	Background on Fixation 2	Label
a)	Consistent	Consistent	BC/BC
b)	Consistent	Grid	BC/Gr
c)	Grid	Consistent	Gr/BC
d)	Grid	Grid	Gr/Gr
e)	Inconsistent	Inconsistent	BI/BI
f)	Inconsistent	Grid	BI/Gr
g)	Grid	Inconsistent	Gr/BI
h)	Grid	Grid	Gr/Gr

The only difference between conditions d and h (both labelled Gr/Gr) was that the objects were located in somewhat different positions within the two-dimensional frame of the scene. If performance differs in these two conditions it must be due to this location difference since they have exactly the same grid background. As a result of having this measure of object location differences, comparisions could be made directly between the Consistent Background conditions and the Inconsistent Background conditions, taking into account the effect of target location.

As in Experiment 2, the same 48 target objects were cued in each of the eight conditions for each subject. This resulted in 384 trials over the entire experiment. These 384 trials were divided into eight blocks of 48 trials. It generally took four one hour sessions for a subject to complete all eight blocks.

The only other difference in design from Experiment 2 was that trials were no longer completely randomized across the experiment. Instead, the order of trials was counterbalanced in the following manner: Each block of 48 trials consisted of one exposure for each target. The condition in which this target appeared was counterbalanced across subjects. Within the block of 48 trials the order was random. Thus, a practice analysis,comparing the first half of the experiment with the second half, would no longer be confounded by having different sets of items in the two halves.

Apparatus and Procedure

The apparatus and procedure were exactly the same as in Experiment 2.

Results and Discussion

Naming Time

The purpose of Experiment 3 was to determine whether background information is utilized from only the first fixation on a scene or from both fixations on the scene. The primary measure was again object naming time. The data shown in Table 5 is the object naming time in Experiment 3 measured from when the target object was fixated. Criteria for including a trial were exactly the same as in Experiment 2. The subject had to a) be fixating the center region of the scene when the trial began (within 1 degree of the center) and b) eventually fixate within one degree of the target object before naming it. Also, as before, trials where a naming error occurred were also not included. Naming errors happened very infrequently, less than 1% of the time.

Table 8

Mean Naming Time in Experiment 3 (in milliseconds)

		Background on 2nd Fixation	
		Consistent Background (BC)	Grid (Gr)
Background on 1st Fixation	Consistent Background (BC)	722	720
	Grid (Gr)	733	741

		Background on 2nd Fixation	
		Inconsistent Background (BI)	Grid (Gr)
Background on 1st Fixation	Inconsistent Background (BI)	790	766
	Grid (Gr)	764	741

Note - Naming time measured from when target object was fixated until target object was named.

The average overall naming time in Experiment 3 was 747 ms. As can be seen in Table 8 naming time in the control conditions (both labelled Gr/Gr) was exactly the same (741 ms). This indicated that there was no effect of the spatial location of the target objects. Therefore it was reasonable to make the direct comparisions between the cells in the top of the table, the Consistent conditions and the cells in the bottom of the table, the Inconsistent conditions. The difference between the Consistent Background (BC/BC) and

Inconsistent Background (BI/BI) conditions was -68 ms. Objects were named faster when they were located in the conistent backgrounds than when they were in the inconsistent backgrounds. This difference was significant, t(7) = 7.4, p < .005, (s = 26). This effect replicated the findings of Experiments 1 and 2.

To evaluate the effect of background information on the first fixation on the scene, naming times can be compared in the BC/Gr condition against naming time in the BI/Gr condition. Subjects were 46 ms faster to name the target object when a consistent background was present on the first fixation than if an inconsistent background was present on the first fixation. This difference was significant t(7) = 4.3, p < .005, (s = 30).

The comparision between the Gr/BC condition and the Gr/BI condition is one indication of the role of background information on the second fixation. Object located in consistent backgrounds on the second fixation only were named 31 ms faster than object located in an inconsistent background on the second fixation only. This difference was significant, t(7) = 3.2, p < .025, (s = 27). Thus, the results of this set of comparisons indicated that a) background information affected object identification most when present for the entire trial, b) background information can be acquired from the first fixation on the scene and c) background information was also acquired during the second fixation on the scene.

Even though there was no difference between the two Grid Background Controls an argument could be made that these conditions should still be included in the comparisons. For this reason, somewhat more complicated comparisons were performed on the data. Since the difference between the grid conditions can be taken as a measure of the effect of object location it would be reasonable to assess the effects of interest by comparing the consistent and inconsistent conditions minus any effect of the object location. Specifically, if one wanted to assess the role of background information on the first fixation the appropriate computation would be (BC/Gr - BI/Gr) - (Gr/Gr - Gr/Gr). The first term in the equation represents the direct comparison between the consistent and inconsistent conditions and the second term represents the difference due to object location. The results of this comparison were very like those reported above. The effect of background consistency when the background was present for the entire trial, (BC/BC - BI/BI) - (Gr/Gr - Gr/Gr), was significant, $t(7) = 6.5$, $p < .005$. The effect of background on the first fixation of the scene, (BC/Gr - BI/Gr) - (Gr/Gr - Gr/Gr), was also significant, $t(7) = 2.5$, $p < .025$. Finally, the effect of background on the second fixation, (Gr/BC - Gr/BI) - (Gr/Gr - Gr/Gr), was also significant, $t(7) = 2.1$, $p < .05$.

Table 9

Mean Differences between Consistent and Inconsistent Conditions in Experiment 3 (in milliseconds)

		Background on Second Fixation	
		Meaningful	Grid
Background on First Fixation	Meaningful	-68	-46
	Grid	-31	0
	Mean	-50	-23

Table 9 shows the differences between the means of Consistent and Inconsistent conditions shown in Table 8 (i.e. the efffects discussed above). The upper left entry in Table 9 (-68 ms) is the difference between BC/BC and BI/BI. Similarly, the upper right entry (-46 ms) is the difference between BC/Gr - BI/Gr, etc. Another way to test the effect of background on the second fixation would be to test for differences in the marginals in Table 9. This test would indicate the difference between the average consistency effect with meaningful backgrounds and the average consistency effect with grid backgrounds on the second fixation. Another way to represent this comparison is [[(BC/BC-BI/BI) + Gr/BC-Gr/BI]/2] - [(BC/Gr-BI/Gr) + (Gr/Gr-Gr/Gr)]/2], where the first term in brackets is the average consistency effect in the conditions where a meaningful background was present on fixation 2 and the second bracketed term represents the average consistency effect when a grid was present on fixation 2. This difference, a 27 ms advantage for meaningful backgrounds, was significant, $t(7) = 3.67$, $p < .01$. Thus, overall there was an effect of background type (either meaningful or grid) on the second fixation on the scene. By comparing the upper two entries in Table 9, one can assess the role of the second background given that a meaningful background was present on the first fixation. This comparison, (BC/BC - BI/BI) - (BC/Gr - BI/Gr), was marginally significant, $t = 1.52$, $.05 < p < .10$. The comparison between the lower two entries on Table 9, a difference of 31 ms, tests the effect of the second background type given a grid on the first fixation. This comparison has already been reported above, and was significant, $t(7) = 2.1$, $p < .05$. Thus, overall there was a significant effect of background type on the second fixation (27 ms); a marginal effect if the background on the first was meaningful (22 ms) and a significant effect if the background on the first was meaningless (31 ms). Furthermore, the difference between this 22 ms effect and the 31 ms effect was not significant.

These results, taken together, indicate that scene backgrounds on the second fixation do affect object identification. The above analyses demonstrated that, although the effect of second background was slightly less given that a grid was presented on the first fixation, this was not significantly different than if a grid had been presented on the first fixation.

The above analyses all indicated that the background affects processing on both fixations on the scene. In fact, the data appear roughly additive. The effect of background information during the entire trial (68 ms) is almost the sum of the effect of background information on the first fixation (46 ms) and the effect of background information on the second fixation (31 ms).

The fact that background information was acquired during the first fixation on the scene is not a surprising result. Context effects have been found when the scene has been presented only for brief durations (see Boyce et al., 1989; Biederman et al, 1982). Hence, it was known that processing of background information could occur very quickly.

The more surprising finding in the current experiment was the effect of background information on the second fixation. Loftus (1983; see also Loftus and Mackworth, 1978) has proposed that during the first fixation on a scene the contextual information is processed and that later fixations on the scene are specifically for the purpose of identifying the objects. This model invokes some sort of attentional mechanism that, metaphorically, could be thought of as a spotlight. During the first fixation on the scene, the attentional spotlight is quite large, encompassing most of the scene. During this time large-grained visual information (perhaps corresponding to the scene background) is acquired. After the context is identified, the eyes move to an object located in the scene. At this time the spotlight narrows to focus all resources on identifying the fixated object.

The results from this experiment indicate that this model is not correct if it assumes that scene background information is not processed after the first fixation. In fact, this experiment indicates that scene backgrounds must be identified at some level on the first fixation (or there should not have been context effects on fixation 1) and that further processing of the background continues during the second fixation. Perhaps the background is identified at a very global level on the first fixation and then further background processing results in a more detailed representation of the background.

This result (the effect of background on fixation 2) also adds to our knowledge about information acquisition and eye movements during scene viewing. First, it can be concluded that the region of the scene that is fixated is not necessarily the only region that is being processed. That is, even while fixating the target object, subjects must have been acquiring additional information about the background. Since the subjects were instructed to name the target object as fast as possible, however, one would think that the best strategy would be to focus all resources on the target object once it has been located and ignore the background. The fact that subjects accumulated background information from the second fixation on the scene indicates that this strategy was not used.

Second, this result suggests that information about the context gained on one fixation on the scene can be carried across an intervening saccade. One would not have expected to get an additional benefit from having the context present for the entire trial if this were not the case. However, this experiment can make no claim about the nature of the code in which information is integrated across subsequent fixations (i.e. whether it is stored in a visual or semantic code).

To summarize, the findings of this experiment replicate the previously found context effect. Furthermore they indicate that background information is acquired during both fixations on the scene. These results will be discussed further in the general discussion.

Eye Movement Latency

A second dependent measure that was included in the data record was subjects' time to respond to the wiggle by moving their eyes off the center of the scene. The mean eye movement latencies for each condition can be seen in Table 10. Averaging across conditions, the mean eye movement

Table 10

Mean Eye Movement Latency in Experiment 3 (in milliseconds)

		Background on 2nd Fixation	
		Consistent Background (BC)	Grid (Gr)
Background on 1st Fixation	Consistent Background (BC)	311	307
	Grid (Gr)	288	290

		Background on 2nd Fixation	
		Inconsistent Background (BI)	Grid (Gr)
Background on 1st Fixation	Inconsistent Background (BI)	312	311
	Grid (Gr)	284	285

latency was 299 ms. A three factor analysis of variance conducted on this data indicated that there was only a main effect of background type on the first fixation F(1,7) = 268.3, p < .00002). That is, if the background on the first fixation was a grid subjects moved their eyes to the wiggled object faster than if the background on the first fixation was meaningful.

One explanation for this result is that since the visual information in the grid background condition was considerably less complex than the meaningful background conditions, there was less information to process on the first fixation when the grid background was present. This, in turn, would allow subjects to register the wiggle faster.

Practice Analysis

As with the previous two experiments an analysis of the effect of practice was conducted. In this analysis, as before, the data from the first four trial blocks constitutes the first half of the experiment and data from the last four trial blocks makes up the second half of the experiment. Because of the counterbalancing of item and order in this experiment, however, practice effects are no longer confounded with item effects. The results from a four way analysis of variance indicated an 85 ms effect of practice ($F(1,7)= 191.6$, $p < .001$). No other factors interacted with practice. Thus, the effects in the second half of the experiment were exactly the same as in the first half.

Items Analysis

The purpose of doing an items analysis on the data was to determine if the effects reported above in the naming time section are the same if one collapses across subject and analyzes by item. Specifically, it was of interest whether or not the three important comparisons (BC/BC vs. BI/BI, BC/Gr vs. BI/Gr, and Gr/BC vs. Gr/BI) would be significant across items. That is, were the observed effects due to a subset of the items included in the experiment or are the effects reliable for all items? Simple effects t-tests were conducted on the naming time data reported in the above section, but collapsed across subject, looking at items. There were significant differences between the BC/BC condition and the BI/BI condition, $t(47) = 4.3$, $p < .005$ and between BC/Gr and BI/Gr conditions, $t(47) = 2.64$, $p < .025$. However, the effect of background on the second fixation (Gr/BC vs. Gr/BI) was only marginally significant, $t(47) = 1.82$, $p < .10$. Thus, the principal effects in the experiment were quite consistent over the various target objects.

In summary, Experiment 3 answered questions about the time course of background information acquisition during scene viewing. Specifically, background information was acquired on both the first and second fixations on a scene. The ramifications of the findings for models of scene perception will be discussed in the General Discussion.

CHAPTER 5

GENERAL DISCUSSION

Introduction

The three experiments reported here indicated that a) context facilitates the object identification process, b) scene backgrounds are one locus of these context effects, c) background information is acquired during the first fixation on a scene and d) background information is also acquired during the second fixation on a scene. The results of these experiments lay to rest the concerns that the prior research using brief exposures overestimated the degree to which context affects object identification. The present "wiggle" paradigm allowed subjects ample time to view the target object and context effects were still found. In addition, the results of Experiment 3 indicate that a strictly serial model of scene processing (as proposed by Loftus, 1983) is also not correct. Background information was acquired on both fixations on the scene. These results have ramifications for models of scene processing and suggest the outline of a model of scene processing that addresses the issues of: a) the mechanism of the facilitation; b) time course of contextual information acquisition; and c) the role of eye movements in scene viewing.

Mechanism of Facilitation

As was outlined in the introduction, there are at least two possible mechanisms of facilitation: Object to Object Priming and Schema Activation. Each of these mechanisms will be discussed in light of the results of the current experiments.

Object to Object Priming

The object to object priming hypothesis is that when related objects occur in scenes together, the identification of one object facilitates the identification of others. There are two problems with the object to object priming hypothesis with regard to this data: a) the formal semantic nature of the priming relationships and b) the magnitude of the priming effect.

This hypothesis views the important relationships between objects to be semantic or associative. That is, "dog" and "cat" are related by nature of their common association with one another and "apple" and "orange" are related because they belong to the same semantic category. I have argued in the introduction that this definition of relations is not sufficient to account for scene context effects. That is, objects that typically co-occur in scenes may not be semantically or associatively related to one another. A glance at the Appendix where the objects in the scenes were listed indicates that most of the objects used in the current experiments are not strongly related to one another under this defininition of relationship. Most importantly, Boyce et al. (1989) Experiment 2 investigated the influence of cohort objects on object identification. Contrary to the object to object priming hypothesis, the cohorts had no effect.

The magnitude of the object to object priming effect found by Henderson et al. (1989) also calls this mechanism into question. In a naming time experiment similar to the current experiments (except without scene context) objects were named 13 ms faster if a related object was fixated prior to the target than if an unrelated object was fixated first. Furthermore, this priming effect was found predominantly when the subjects were given no preview of the target object. In contrast, across the current three experiments the average effect of background consistency was 64 ms. That is, objects were named 64 ms faster in the consistent backgrounds than in the inconsistent backgrounds. Thus, even if one could demonstrate that an object to object priming mechanism is at work during <u>scene</u> processing, it would have a very small effect relative to the effect of scene backgrounds on object identification. Thus, the bulk of the evidence suggests that object idenfication is facilitated by scene backgrounds and that the simple object to object priming explanation proposed by Henderson et al. cannot account for these findings.

However, a more sophisticated version of the object to object priming hypothesis may be able to account for this data. If one assumes that backgrounds are large objects and that the associations between objects can be episodic as well as semantic then the object to object priming hypothesis may be able to account for this data. However, this model still leaves the lack of cohort effects to be explained. That is, this model would have to also assume some greater strength of relationship between the backgrounds and objects than between the objects that co-occur in the scenes.

Schema Activation

A schema activation account of the present paradigm is as follows: Identification of the background leads to the activation of the appropriate schema for the scene. Once this superordinate node in the schema is activated, it facilitates the identification of objects likely to fit into the context; that is, object representations that are connected to this superordinate scene node.

While this account seems reasonable, it raises the question of how this schema activation actually facilitates object identification. It seems as though there are two ways that this facilitation can occur: a) background identification could speed up further perceptual processing or b) background indentification could limit the amount of further perceptual processing necessary. Each of these possible mechanisms will be discussed.

We know that the rate of visual information acquisition is not constant. Many things can alter the rate at which one can extract visual features. For example, Loftus (1985; see also Loftus, Nelson and Kallman, 1983) has found that the intensity of the stimulus display is closely related to the speed of acquisition of pictorial information. T-scope experiments with letters show that perceptual information is acquired slower from the periphery than from the fovea (Estes, 1978; Sperling, 1970). Perhaps context affects object identification by speeding up the rate at which visual information is acquired. That is, once the appropriate schema has been activated, perceptual processes speed up and object information is accumulated at a faster rate. The largest problem with this mechanism is that it is not clear how the higher order processes can "feed back" and alter the mechanics of perception. For example, it is hard to imagine that higher order processes affect the speed at which photochemicals in the retina work. Perhaps the feature acquisition is somehow speeded in the primary visual cortex, but it is still not clear how semantic information can be directed back to affect these processes.

It seems more plausible that scene context limits the amount of perceptual processing necessary for identification to occur. That is, when the appropriate schema is activated in memory, activation is spread to the related subordinate object nodes. If one assumes that there is some threshold that must be reached for object identification to occur, one could think of this additional activation to the object node as equivalent to lowering its threshold. In either case, less input from the perceptual system is necessary for threshold to be reached. This model does not assume context modulates the rate at which perceptual information accrues. For example, when the kitchen background is identified, activating the "kitchen node" in memory, activation is spread to the related concepts "toaster", "glass", etc. When visual information about a toaster is input into the system it takes less time for the toaster node in memory to reach threshold. Thus, the name TOASTER can be said quickly. This mechanism is similar to that proposed by McClellend and Rumelhart (1981). The important aspect of this explanation is that it does not require that perceptual processes are altered by context, only that one can "act" earlier on the basis of less advanced perceptual processing.

This mechanism would also predict that there would be little or no interference in the Inconsistent Background condition. If the background is identified as a bathroom, instead of a kitchen, one would expect that toaster would not be facilitated, since bathroom is not connected to toaster in memory and one would not expect this to interefere. In the inconsistent background scenes, object representations were still facilitated by the context, it was just not the objects that were pictured in the scene. That is, in the bathroom background with the kitchen objects (Inconsistent Condition) activation from bathroom node would have spread activation to plunger, toilet paper, etc. This activation would not have affected speed of recognition for toaster which was not facilitated or interfered with by the identification of the bathroom background. The results of Boyce et al. (1989) Experiment 2 would support this hypothesis. It was found that the only predictor of context effects was the relationship between the target object and the scene background.

There is one result from the current experiments that does not fit very well with this schema explanation. In Experiment 1, the scene consistency effect was not highly correlated with ratings of how well the objects fit in the scene context. An independent group of subjects rated the objects for their degree of consistency with the backgrounds and the correlation of this measure with the difference between identification in the Consistent and Inconsistent conditions was only 0.2. If the schema view was correct then one would have expected this correlation to be higher. That is, the more predictable the object was from the background, the more activated that object node should have been. This is a somewaht puzzling result in light of the manner in which these scenes were created. In order to generate the ideas for the scenes, an independent group of subjects listed objects that they thought were most likely to occur in the backgrounds. This is the typical way of examining the contents of peoples' schemas (see Mandler, 1984).

This result does not necessarily invalidate a schema organization, but it may suggest that schema activation is a more interactive process. That is, it is possible that the scene background provides initial access to a general schema for the context, such as "room", or even "bedroom", and at the same time partial identification of the target object facilitates another, more specific schema, for example, "child's room". This activation spreading down from a more global scene schema and up through partial identification of the target object may result in greater activation for the more specific schema. The target object may be highly predictable with respect to this more specific schema, but only plausible in the more generic "bedroom" schema. Since the ratings were obtained by giving subjects generic labels for the scenes, such as "bedroom scene", the moderate correlation may not truly reflect the degree to which the objects were <u>predictable</u> from the background.

The organization of schemas is still not well understood and their proponents have been widely criticized because schema theory makes few predictions and because when it is adapted to account for a wide range of phenomena, many ad-hoc (and potentially inconsistent) assumptions are made. I do agree with some of these criticisms. However, the results of these experiments, taken together with the previous literature on context effects (Antes, 1977; Antes et al., 1981; Biederman, 1972; Biederman, Glass and Stacy, 1973; Biederman et al. 1982; Boyce et al, 1989; Friedman, 1979; Loftus et al., 1978), indicate that the relationship between context and the object identification process is not a simple associative one.

No modular semantic network explanation is going to be able to account for these findings, to the extent that the relationships in the semantic network are not based on the probability of co-occurence. The modified version of the object to object priming hypothesis, which begins to resemble the schema organization, may be able to account for the data because backgrounds may function in the network as large objects. However, the lack of a cohort effect indicates that the relationship between backgrounds and the objects likely to occur in them is stronger than the relationship between the scene objects themselves. Even the modified object to object priming network cannot account for this. The hierarchical organization of schemas can account for these effects. A schema organization that posited strong connections between the background and the objects and only weak, or non-existent connections between the objects best accounts for the data.

Time Course of Context Information Acquisition

The major finding with respect to the time course of contextual processing was that consistent backgrounds facilitated object identification on both fixations on the scene and furthermore, this effect was roughly additive. This is a somewhat puzzling result if one thinks that identification of context is an all or nothing process. That is, how could it be that the background present on the first fixation was identified enough to facilitate object identification but then further information extracted on the second fixation led to greater facilitation? There are at least two mechanisms that could account for the additive effect of background on both fixations. The first I will call the "levels of representation" hypothesis and the second the "neighborhood" hypothesis.

The levels of representation hypothesis holds that backgrounds are represented in memory at various levels of specificity. Perhaps there is a node superordinate to "kitchen" that represents something like "room in a house". Or, maybe there is a level between "kitchen" and "room in a house" that we do not have a word for in our language. The idea here is that there is a more general representation for the backgrounds' meaning stored in memory. During the first fixation on the scene, it is this level of representation that is activated. This activation partially facilicitates the representations for the objects that could occur in this scene. During the second fixation on the scene, the background is processed further and this results in contact with the more detailed and specific representation for the background. This also facilitates object identification. Thus, this hypothesis would argue that there are multiple levels of representations stored in memory for each scene and the facilitation for the target object was two sources; the more general representation and the more specific representation.

A second mechanism is the "neighborhood" hypothesis. Background information acquired on the first fixation on the scene could activate more than one schema. Backgrounds that share many of the same visual characteristics could be activated by the information acquired on the first fixation. This would weakly activate objects that are consistent with the scene background and objects that are consistent with visually similar backgrounds. During the second fixation on the scene, additional information about the background is acquired and this raises the activation for one schema above that of other schemas. At this point, one interpretation of the background has been arrived at and this results in further facilitation for only the objects that are consistent with this background. This hypothesis holds that backgrounds and their related objects are only represented once in memory.

The major difference between the two models is the following: in the first model only a subset of the objects eventually activated are activated early in the processing. In contrast, the second model posits that a larger number of objects (including inappropriate ones) are activated early in processing and then are narrowed down further by additional processing of the background. The data from the current experiments cannot distinguish between these two mechanisms. However, they make testable predictions. If the "neighborhood" mechanism is correct, than having background information present only on the first fixation on the scene should result in facilitation for objects that are unrelated to the current background (but related to a visually similar background) as well as facilitating the correct set of objects.

One interesting aspect of the results of Experiment 3 is that it appears as though processing of the background information past the first fixation on the scene is obligatory. That is, even when the target object had been located and the subject had fixated the target, background information acquisition continued. This result would argue for obligatory processing of background to some level of comprehension. Thus, it appears as though background information is so important to the process of object identification that the extraction of meaning from the background continues even while the target object is being identified. In the levels of processing hypothesis the further background processing results in a contact with a more detailed schema for the scene. According to the neighborhood hypothesis, this further background processing on the second fixation results in the elimination of the activation for visually similar background schemas.

Eye Movements, Covert Attention and Scene Processing

The primary purpose of the current research was to investigate questions about the role of context in object identification. However, the results of the current experiments do add to our knowledge about eye movements, covert attention and information acquisition in scene viewing. Specifically, these experiments speak to the questions of extrafoveal information acquisition and integration of information across saccades.

Perceptual Span and Scene Perception

"Perceptual span" refers to the region of text from which a reader is acquiring useful information (Rayner, 1975). Rayner (1975) determined that in addition to acquiring information from the word that is in the fovea, information to the right of the currently fixated word can be acquired as well. The quality of this extrafoveal information varies as a function of its distance from the fixation point.

Some research has been conducted to determine the size of the perceptual span in scenes. Saida and Ikeda (1979), using the moving window technique, concluded that the perceptual span for scenes was about 50% of the area surrounding the fixation point. That is, on a given fixation, useful information was being acquired from about 50% of the scene. Nelson and Loftus (1980) found that little information about details in the scene a fixation could be extracted from further than about 1.5 degrees from fixation.

The current experiments indicated that extrafoveal information can be extracted from a large region of the scene. In all of the experiments, the first fixation on the scene was in the center. These scenes were designed so that very little useful background information would be provided foveally when fixating the center of these scenes and there was never an object in the center of the scene. Thus, any background information that was acquired during the first fixation must have been acquired from extrafoveal processing. The results from Experiment 3 indicated that context information was acquired during the first fixation on the scene.

The fact that context information was acquired extrafoveally during the first fixation on the scene might not be too surprising in light of the fact that there was very little foveal processing to be done in the center of these scenes. Hence, during the first fixation on the scene all of one's resources could be devoted completely to peripheral information extraction. The more surprising result was the effect of context on the second fixation. Extrafoveal information was processed even when the target object was in the fovea.

These data do not indicate the exact size of the perceptual span but they do indicate that it is larger than 2 degrees. The visual information that conveyed the meaning of the backgrounds was spread over a large region. In a separate pilot study, subjects were not able to accurately identify the background when given only the center 2 degrees of the scene. Thus, background information acquired on the first fixation must have been from a region larger than 2 degrees. Similarly, since backgrounds were processed during the second fixation on the scene and we know that the target object (approximately 2 degrees in size) resided in the fovea at this time, we can conclude that the perceptual span is considerably larger than two degrees.

In summary, from these results it can be concluded that a) the size of the perceptual span is large (at the very least larger than 2 degrees) and b) the size of the perceptual span does not vary drastically from the first to the second fixation on the scene.

The fact that background information was acquired during the fixation on the target object indicates that we should use caution when employing a measure like fixation duration as a measure of object identification time. Since more than the target object was being processed during the second fixation on the scene, therefore, fixation durations are likely to reflect not only time to identify the object, but other factors as well, such as processing of the context.

Integration of Information Across Saccades

Some kind of information is integrated across saccades, but the major unresolved question is what kind (level) of information is undergoing the integration. The bulk of the evidence from experiments employing text indicates it is not basic "retinal" information that is integrated but something more abstract (Rayner, McConkie and Zola 1980; McConkie and Zola, 1979). Similar experiments with pictures (Pollatsek, Rayner and Collins, 1984) found that some sort of abstract shape information was also integrated across saccades.

The current experiments indicate that the background information acquired on one fixation can affect processing on a subsequent fixation. The major evidence for this comes from the added advantage of having a consistent scene background present on both fixations over having it present on just one of the fixations. Since the advantage of having a consistent background on both fixations was substantially greater than having the background present only on the second fixation on the scene, background information was integrated across saccades.

The results of Experiment 3 also suggest that the information integrated across the saccade was not "featural" information but some higher order code. If integration was attempted with some visual feature level information one might expect to get large disruption effects in the conditions where major changes in the background information occurred from fixation to fixation. The average naming time in the four background switching conditions was 746 ms compared with 749 ms where much smaller changes occurred (only a size change). Thus, it appears as though something more conceptual (i.e. the meaning of the background) was integrated across successive fixations on the scene. In summary, the results of the current experiments corroborate the view that higher order conceptual information is integrated across saccades.

Allocation of Covert Attention

There at least two ways in which attention could be allocated during scene processing. The first, which I have already discussed, could be termed the "zoom model". In this view, during the first fixation on a scene, attention is diffused over the entire visual field. Subsequently, the spotlight of attention is narrowed to encompass only the target object. The second model of attention allocation is one proposed by Henderson, Pollatsek and Rayner (1989) which they term the sequential attention hypothesis. This model states that attention is first allocated to the foveal region, and then after some threshold of identification of foveal information has been crossed, attention is shifted to a new region of the visual field.

The results of Experiment 3 seem to rule out the "zoom model" as proposed by Loftus. It was not the case that the background information was attended to only during the first fixation on the scene. This result is somewhat suprising, given that the zoom strategy seems like an efficient way to process scene information. Since enough background information was extracted during the first fixation on the scene to cause facilitation for the target object, it seems like it would have been a good strategy to focus all processing resources on the target object once is had been located. This is not what happened.

Henderson et al. (1989) also tested the valdity of the zoom hypothesis. In their experiments (done with arrays of objects and not scenes) they found that attention was not first allocated to all regions of the display, but instead was allocated in a sequential manner. That is, attention was first allocated to the foveal object and then allocated to the object that was to be fixated next. They suggested that the allocation of attention during scene viewing may function in the same fashion.

There is some weak evidence in the current experiments against this sequential attention mechanism. In order to account for the context effects on the second fixation on the scene with this mechanism, one would have to propose that processing of the foveal target object was done first and then attention was allocated to some other region of the scene (perhaps somewhere on the background). It is not clear that this could account for the results obtained. First, it would not be effecient to shift attention off the target object in the current paradigm once it had been located. Second, it is unclear how context could have an effect during the second fixation if the background was attended to only after the target object (foveal region) had been processed. Thus, this serial mechanism does not seem to be at work in the current paradigm. Instead it appears as though during the first two fixations on a scene, attention is allocated to the entire visual field and this does not change over the course of the trial. The evidence from the current experiments thus indicates that neither simple mechanism -- zoom or sequential shift -- is an adequate model of selective attention during the first few fixations on a scene.

Conclusion

I have outlined a model of scene perception which accounts for the results of the three experiments reported in this dissertation. The model suggests that the identification of scene backgrounds activates a schema for the scene stored in memory. Once the top node in the correct schema has been activated (this node represents the background) the activation spreads to objects that are consistent with the background. Because the object nodes consistent with the background have been partially activated, less perceptual information is necessary for the identification threshold to be reached.

The process of contacting the correct schema is not necessarily completed in the first fixation. Partial identification of background information might either a) make contact with some superordinate, less specified, representation initially, or b) activate a neighborhood of related schemas. Further background processing during the second fixation on the scene results in full activation of the correct schema. Also, I have argued that the processing of background information is obligatory until some criterion of background identification has been reached.

Finally, it seems that the perceptual span for pictures is large and that extrafoveal information plays an important role in scene processing. Moreover, the results of

Experiment 3 suggest that conceptual information from the background is integrated across saccades.

APPENDIX

DESCRIPTIONS OF SCENES

Scene name in parentheses indicates scene that was used for background inconsistent conditions.

* Indicates non-target object in all experiments
** Indicates non-target object in Experiments 2 and 3.

<u>Scenes and Objects</u>

1. Bedroom
   (Refrigerator)

   1. Teddy Bear **
   2. Doll
   3. Suitcase
   4. Cap
   5. Pennent *

2. Broom Closet
   (Desk)

   1. Iron **
   2. Scrub Brush
   3. Paper Towels
   4. Bucket
   5. Wisk Broom *

3. Clothes Closet
   (Oven)

   1. Pants **
   2. Bowtie *
   3. Glove
   4. Shoe
   5. Hat

4. Construction
   (Porch)

   1. Drill
   2. Hammer **
   3. Saw
   4. Saw Horse *
   5. Ladder

5. Desk
   (Broom Closet)

   1. Briefcase
   2. Lamp
   3. Phone
   4. Stapler **
   5. Picture *

6. Diner
   (Shower)

   1. Cup
   2. Fork
   3. Salt **
   4. Syrup
   5. Ketchup *

7. Fireplace
   (Laundry)

   1. Candle
   2. Chair **
   3. Clock
   4. Logs
   5. Bellows *

8. Refrigerator
   (Bedroom)

   1. Butter
   2. Cheese
   3. Lettuce *
   4. Milk
   5. Onion **

9. Laundry
   (Fireplace)

   1. Basket
   2. Laundry Soap Box *
   3. Hanger
   4. Shirt
   5. Bleach **

10. Oven
    (Clothes Closet)

    1. Pot
    2. Spoon **
    3. Teapot
    4. Turkey
    5. Oven Mit *

11. Pool
(Street)

1. Flipper
2. Raft
3. Grill
4. Ball **
5. Life Saver *

12. Porch
(Construction)

1. Birdhouse **
2. Skate
3. Pumpkin
4. Newspaper
5. Watering Can *

13. Kitchen Sink
(Toilet)

1. Coffee Maker
2. Eggbeater **
3. Toaster
4. Wine Glass
5. Dish Detergent *

14. Shower
(Diner)

1. Bath Mat **
2. Slippers
3. Soap *
4. Towel
5. Shampoo

15. Street
(Pool)

1. Bike
2. Fire Hydrant
3. Mailbox
4. Parking Meter *
5. Wagon **

16. Toilet
(Kitchen Sink)

1. Toilet Brush **
2. Toilet Paper
3. Kleenex
4. Toilet Plunger
5. Baby Powder *

REFERENCES

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

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

Antes, J. R., and Man, S. W. (1984). Global-local precedence in picture processing. Psychological Research, 46, 247-259.

Antes, J. R., Man, S. W., and Penland, J. G. (1981). Local precedence in picture naming. the importance of obligatory objects. Presentation at 1981 meeting of the Psychonomic Society.

Antes, J. R., and Penland, J. G. (1981). Picture context effects on eye movement patterns. In Fisher, Monty and Senders (Eds.), Eye Movements: Cognition and Visual Perception, Hillsdale, NJ, Earlbaum.

Antes, J. R., Penland, J. G., and Metzger, R. L. (1981). Processing global information from briefly presented pictures. Psychological Research, 43, 277-292.

Biederman, I. (1972). Perceiving real-world scenes. Science, 177, 77-80.

Biederman, I. (1981). On the semantics of a glance at a scene. In Kubovy and Pomerantz (Eds.), Perceptual Organization, Hillsdale, NJ, Erlbaum.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. Psychological Review, 94, 115-147.

Biederman, I., and Blickle, T. W. (1985). The perception of objects with deleted contours. Unpublished manuscript, State University of New York at Buffalo.

Biederman, I., and Ju, G. (1988). Surface versus edge-based determinants of visual recognition. Cognitive Psychology, 20, 38-64.

Biederman, I., Mezzanotte, R. J., and Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing violation. Cognitive Psychology, 14, 143-177.

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

Blanchard, H. E. (1985). A comparision of some processing time measures based on eye movements. Acta Psychologica, 58, 1-15.

Boyce, S. J., Pollatsek, A., and Rayner, K. (1988). Role of background information on object identification. Journal of Experimental Psychology: Human Perception and Performance, 15, 556-566.

Estes, W. K. (1978). Perceptual processing in letter recognition and reading. Handbook of Perception, Vol. IX, New York, Academic Press.

Friedman, A. (1979). Framing pictures: The role of knowledge in automatized encoding and memory from gist. Journal of Experimental Psychology, 3, 316-355.

Henderson, J. M., Pollatsek, A., and Rayner, K. (1987). The effects of foveal priming and extrafoveal preview on object identification. Journal of Experimental Psychology: Human Perception and Performance, 3, 449-463.

Henderson, J. M., Pollatsek, A., and Rayner, K. (1989). Covert visual attention and extrafoveal information use during object identification. Perception & Psychophysics, 45, 196-208.

Kroll, J. F., and Potter, M. C. (1984). Recognizing words, pictures and concepts: A comparison of lexical, object and reality decisions. Journal of Verbal Learning and Verbal Behavior, 23, 39-66.

Loftus, G. R. (1983). Eye Fixations on Text and Scenes. In Rayner (Ed.) Eye Movements in Reading. Academic Press: New York.

Loftus, G. R. (1985). Picture perception: Effects of luminance on available information and information extraction rate. Journal of Experimental Psychology: General, 114, 342-356.

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

Loftus, G. R., Nelson, W. W., and Kallman, H. J. (1983). Differential acquisition rates for different types of information from pictures. Quarterly Journal of Psychology, 35A, 189-198.

Mandler, J. M. (1984). Stories, scripts and scenes: Aspects of schema theory. Hillsdale, NJ: Earlbaum.

McClelland, J. L., and Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. Psychological Review, 88, 375-407.

McConkie, G. W., and Zola, D. E. (1979). Is visual information integrated across successive fixations in reading. Perception and Psychophysics, 25, 221-224.

Metzger, R. L., and Antes, J. R. (1983). The nature of processing early picture perception. Psychological Review, 45, 267-274.

Nelson, W. W., and Loftus, G. R. (1980). Functional visual field during picture viewing. Journal of Experimental Psychology: Human Perception and Performance, 6, 391-399.

Ostergaard, A. L., and Davidoff, J. B. (1985). Some effects of color on naming and recognition of objects. Journal of Experimental Psychology: Learning, Memory and Cognition, 11, 579-587.

Pollatsek, A., Rayner, K., and Collins, W. (1984). Integrating pictorial information across eye movements. Journal of Experimental Psychology: General, 113, 426-442.

Rayner, K. (1975). The perceptual span and peripheral cues in reading. Cognitive Psychology, 7, 65-81.

Rayner, K., and Pollatsek, A. (1987). Eye movements in reading: A tutorial review. In Coltheart (Ed.), Attention and Performance XII, Erlbaum.

Rayner, K., McConkie, G. W., and Zola, D. E. (1980). Integrating information across eye movements. Cognitive Psychology, 12, 206-226.

Rock , I. (1983). The Logic of Perception, Cambridge Ma., MIT Press, Bradford Books.

Saida, S., and Ikeda, M. (1979). Useful field size for pattern perception. Perception and Psychophysics, 25, 119-125.

Selfridge O. G., & Neisser, U. (1960). Pattern Recognition by machine. Scientific American, 203, 60-68.

Sperling, G. (1970). Short-term memory, long-term memory and scanning in the processing of visual information. In Young, F. A. & Lindsley, D. B. (Eds.) Early Experience and Visual Information Processing in Perceptual and Reading Disorders. Washington D.C.: National Academy of Sciences.

Theios, J., and Amrhein, P. C. (1988). A theoretical analysis of the cognitive processing of lexical and pictorial stimuli: Reading, naming and visual and conceptual comparisons. Psychological Review, 96, 5-24.