Context and object recognition.

John M. Henderson

University of Massachusetts Amherst

Follow this and additional works at: https://scholarworks.umass.edu/theses
CONTEXT AND OBJECT RECOGNITION

A Thesis Presented

by

JOHN MICHAEL HENDERSON

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

MASTER OF SCIENCE

May 1986

Psychology
CONTEXT AND OBJECT RECOGNITION

A Thesis Presented
by
JOHN MICHAEL HENDERSON

Approved as to style and content by:

Alexander Pollatsek, Chairperson of Committee

Rachel Clifton, Member

Keith Rayner, Member

Arnold Well, Member

Seymour Berger, Department Head
Psychology
ACKNOWLEDGEMENTS

I would like to express my appreciation to my committee, Sandy Pollatsek, Keith Rayner, Arnie Well, and Rachel Clifton. I owe special thanks to Keith, for providing helpful comments and discussion all the way through this work. I owe extra-special thanks to Sandy, who has served as an advisor but treated me as a colleague. Sandy asked the hard questions when they needed to be asked, and helped me to always keep a sense of humor by always keeping his.

I would also like to thank my friends in the department for both their intellectual stimulation and their inspired craziness. Dolores Shank, Randy Hansen, Chris Young, Laurel Long, and Pat Collins all earned the right to have their names immortalized here on this page. Peter DeGraef deserves special thanks; I probably spent more time talking with him about the ideas presented here than with anyone else.

Finally, I would like to thank Fernanda Ferreira. She has provided me with both personal and intellectual support. There are other reasons for thanking Fernanda, but she already knows what they are.
This thesis examines the effects of context on object identification. In the past, researchers have argued that the effects of a scene context on object recognition can be explained by postulating that a scene allows activation of a high level memory structure or schema for the scene, which then facilitates the identification of objects which are consistent with that schema. The present work attempts to determine whether an alternative explanation for these context effects is viable. The alternative, called here "intralevel priming", suggests that automatic priming of unidentified objects from already identified, semantically related objects may be able to account for the context effects in scenes. Experiment 1 demonstrates that an object fixated on Fixation N can facilitate the identification of a related object fixated on Fixation N+1. Experiment 2 provides evidence that this effect is due to an automatic rather than a capacity demanding process. In addition,
both experiments show how this intralevel priming mechanism is affected by perceptual factors: The priming effect tends to be larger when the target object can not be easily seen in the parafovea. The results of these experiments are taken to provide evidence that intralevel priming provides a viable alternative to schema theory as an explanation for the context effects found in the visual processing of scenes.
# TABLE OF CONTENTS

Acknowledgements........................................ iii
Abstract.................................................. iv
List of Tables........................................... vii
List of Figures........................................... ix

Chapter

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

  How Does Context Aid Object Identification In Scenes? ....... 2
  Can a Schema be Accessed Quickly Enough to Aid Object Identification? ...... 3
  Evidence for Schema Theory.................................. 11
  Eye Movements and Picture Viewing............................ 25
    Context Effects on First Fixation Duration.................. 26
  General Criticisms and an Alternative Model.................. 38
    Intralevel Priming......................................... 44

II. EXPERIMENTS........................................ 47

  Experiment 1............................................ 48
    Method.................................................. 53
      Subjects.............................................. 53
      Materials............................................ 53
      Apparatus.......................................... 54
      Procedure........................................... 55
    Design.................................................. 56
    Results and Discussion.................................. 59
      Analysis Without Baseline............................... 70
  Experiment 2............................................ 75
    Method.................................................. 77
      Subjects.............................................. 77
      Materials............................................ 77
      Apparatus.......................................... 78
      Procedure........................................... 79
    Design.................................................. 80
    Results and Discussion.................................. 80

III. ADDITIONAL ANALYSES............................... 88

  Priming, Preview, and Eccentricity......................... 88
  Items Analysis.......................................... 94
  Name Frequency Analysis.................................. 94
  Visual Similarity Analysis................................ 96
  Naming Latency Frequency Distribution Analysis.......... 98

vi
IV. DISCUSSION ............................................. 106
   A Model .............................................. 106
   Conclusion .......................................... 114

REFERENCES .............................................. 120
LIST OF TABLES

1. Target objects and their related foveal primes... 56
2. Mean time to name the target object by eccentricity, preview, and prime, Expt 1... 60
3. Mean naming latencies for the three levels of foveal prime and two levels of preview, Expt 1......................... 66
4. Mean naming latencies for the three levels of foveal prime and two levels of eccentricity, Expt 1.......................... 68
5. Mean naming latencies with an object versus a blob in the fovea, Expt 1....................... 69
6. Amount of priming by preview and eccentricity, Expt 1.......................... 71
7. Parafoveal preview effect as a function of eccentricity and foveal stimulus, Expt 1... 73
8. Mean time to name the target object by eccentricity, preview, and prime, Expt 2... 81
9. Amount of priming by preview and eccentricity, Expt 2.......................... 86
10. Priming effect in Experiments 1 and 2 by visual similarity.......................... 97
11. Activation level of a pictogen depending on type of prime and parafoveal preview......... 111
12. Length of time to reach threshold by condition.......................... 112
13. The priming effect as predicted by the model... 113
LIST OF FIGURES

1. The flow of information in a schema model........ 14
2. Naming frequency distributions for 5 degree related trials, preview and no preview.... 102
3. Naming frequency distributions for 5 degree unrelated trials, preview and no preview.. 103
4. Naming frequency distributions for 10 degree related trials, preview and no preview.... 104
5. Naming Frequency distributions for 10 degree unrelated trials, preview and no preview.. 105
CHAPTER I
INTRODUCTION

This thesis examines the effects of context on object identification in pictures. In general, context has been shown to facilitate recognition of a wide variety of perceptual stimuli. For example, letters are more easily identified in words than in nonwords (Reicher, 1969; Wheeler, 1970); words are more easily identified with single word contexts (Meyer, Schvenevedlt, & Ruddy, 1975), single object contexts (Sperber, McCauley, Ragain, & Weil 1979), and sentence contexts (Stanovich & West, 1983); and parafoveal words are more easily identified with constraining semantic contextual information than without (Balota, Pollatsek, & Rayner, 1985; Balota & Rayner, 1983; McClelland & O'Regan, 1981). Using pictures of objects as stimuli, it has been shown that identification is facilitated when an object is presented in a coherent scene (Biederman, 1972) but is inhibited if the object violates its ordinary relation to the visual context (Biederman, 1981); object naming is facilitated by both single object and single word contexts (Kroll & Potter, 1984; Sperber et al., 1979); object misidentification is more likely if the target object visually resembles another object which would be more likely in a given context
(Palmer, 1975); and studies recording eye movements have generally concluded that an object in a semantically appropriate context is more easily identified than an object which does not fit the context as well (Antes, 1974; Friedman, 1979; Loftus & Mackworth, 1978).

How Does Context Aid Object Identification in Scenes?

There are currently two main hypotheses regarding the nature of contextual effects on object identification in scenes. The first approach is to assume that higher level memory representations known as frames (Minsky, 1975), schemas (Bartlett, 1932; Norman & Rumelhart, 1975; Rumelhart, 1980), or scripts (Schank & Abelson, 1975) interact with incoming perceptual information during object identification. On this view, context is facilitative because it acts to invoke the appropriate memory structure (henceforth, schema). Objects which are obligatory in the schema are encoded more or less automatically (with a minimal use of processing resources), while objects which do not fit the schema as well require a more resource expensive encoding process, and objects which do not fit the schema at all require active hypothesis testing which is extremely resource expensive (Friedman, 1979; Friedman & Liebelt, 1981). A second approach, which is at this point a fledgling
hypothesis but which will be investigated in the current paper, suggests that context effects are produced primarily through intralevel priming. On this view, the information available to the object identification stage is perceptual information from lower levels of processing and information about other objects that have already been identified (intralevel information), but not higher level information. This approach may be referred to as the intralevel priming approach and is consistent with the concept of modularity (Fodor, 1983; Marr, 1982).

Can a Schema be Accessed Quickly Enough to Aid Object Identification?

In order for a memory structure such as a schema to effectively aid object identification, the memory structure must be activated and functional prior to (or at least simultaneously with) object identification processing.

Several studies have indicated that the "gist" or general meaning of a scene can be extracted very quickly, in the range of 100 milliseconds (Potter, 1975, 1976; Intraub, 1981). For example, in a seminal study by Potter (1975), subjects viewed a steady sequence of sixteen pictures at a rate of 125 milliseconds per picture. The task was to detect a particular target picture in the
sequence. The target picture was identified for the subject prior to exposure of the sequence either by showing the subject the picture itself, or by giving the subject a brief verbal description of the picture (e.g., "two men drinking beer"). Even at this high rate of exposure, more than 70 percent of the target pictures were detected, regardless of the manner in which the target was originally identified to the subject, though far fewer were remembered in a later recognition test. Potter argued that on the average about 70 percent of the pictures in the sequence must have been identified, since a picture would have to be conceptually identified in order to have been compared with a verbal description. Potter (1976) again found similar results in a replication experiment, when 64 percent of the named targets were detected at a presentation rate of 113 milliseconds per picture.

While Potter does not discuss her findings in terms of schema theory (and perhaps would be comfortable with a more bottom-up interpretation), these results have often been cited as evidence for the view that a general schema for a scene can be accessed in approximately 100 milliseconds, well within the duration of the first fixation in normal picture viewing. The argument put forth by schema theorists (see Biederman, 1981) is that the identification of a scene's general conceptual
meaning is equivalent to the activation of a schema for that scene. Further, according to schema theory, schema activation takes place before many (if any) objects have been identified and acts to guide further encoding of the scene. Since Potter finds rapid encoding of pictures to a semantic level, schema theorists take this as evidence that schemas can be quickly accessed. However, it should be noted that even if subjects are able to quickly encode a picture to a semantic level, this in no way implies that the top-down processes envisioned by schema theorists necessarily aid in the encoding process. (In fact, it is not quite clear what encoding is left to be done once a semantic level representation has been achieved.) In addition, the stimuli used in these experiments were pictures of either a single object or very few objects, and the task was such that identification of a single foveal object would likely be sufficient for a correct response. Therefore, these experiments may have more to say about rapid object identification than about rapid scene identification.

It could also be argued that the above results do not in fact provide evidence for fast semantic activation, especially fast activation without foreknowledge, since Potter always precued subjects with the target picture or a description of the target picture. With precuing, expectancy or set may have
selectively facilitated target identification, for example, by priming object level representations, while having had no effect at all upon non-target processing. This in fact seems quite likely, since the stimuli used in these experiments were designed in a way that would allow subjects to identify the target picture in a sequence through identification of one salient object, with no need to process any other objects contained in the picture (see Potter, 1976, p. 511). If it were the case that expectancy allowed priming of the "cognitive demon" or object level representation for a particular object in the target picture, then the conclusion that a high percentage of non-target pictures had been identified would be unfounded. In fact, the conclusion that even the target pictures had been processed to a semantic level might be mistaken. Since the main issue for the purposes of this paper is whether the meaning of a scene (i.e. a schema) can be activated quickly enough to aid object identification, it becomes of concern whether Potter's results can be accounted for by expectancy. If they can, then the results say little about unprimed schema activation.

One finding seems to directly contradict the expectancy hypothesis. In Experiment 2 of Potter (1976), an explicit prediction of the expectancy hypothesis was tested: Less processing should be needed to reject a
non-target when the target is identified to subjects by the picture itself than when it is identified by a verbal description. In the former case, non-targets could be rejected (and targets accepted) on a purely physical basis, with no need to process them to a semantic level, while in the latter case, a deeper level of processing would be needed. In the original experiments, Potter's cues were the pictures themselves. If her results had been due to perceptual expectancy, then in a recognition memory test, subjects' memory for non-targets should be better after having been precued with the description than after having been shown the picture itself (e.g., Tulving & Gold, 1963). Since recognition memory was found not to differ significantly in the two conditions, the expectancy hypothesis was not supported. These results were taken by Potter to indicate that the non-target pictures in the sequence were processed to a semantic level even when the target could have been picked out on purely physical basis.

However, an alternative explanation is possible. Suppose subjects were identifying the target picture not by analyzing each picture to a semantic level, but by looking for one salient object which should only occur in the target picture. If this were true, then giving either the target picture itself or a verbal description should have the similar effects, namely to
prime a "cognitive demon" for that object. In this case, as Potter (1976) found, target detection would be better given a picture of the target, since a more accurate object level representation could be primed. In addition, it might be expected that encoding of and recognition memory for non-targets in the two conditions would be equal, except that instead of describing the situation as one in which recognition memory in the picture cue condition was "as good as" that in the description cue condition, as one would put it if arguing against the expectancy explanation, it might better be said that recognition memory in the description cue condition was "as poor as" that in the picture cue condition. This is because in both cases, the non-targets are processed at a superficial level, while the targets are recognized by virtue of expectancy, or priming for a particular object. Evidence that this might be a better way of looking at these data is that the recognition rate for non-targets in both conditions was less than 10% correct at a presentation rate of 113 msec per picture and only 20% correct across all presentation rates: 113, 167, 250, and 333 msec.

One other reason that has been given for rejecting the expectancy hypothesis is equivocal. Using the Potter paradigm, Intraub (1981) found that 35 percent of the targets were correctly identified when subjects were
given a negatively phrased description of the target (e.g., "the picture that is not of house furnishings and decorations"). Intraub argued that set or expectancy could not account for these results, since it would be difficult to set up an expectancy for a target based on knowing only what that target is not. Accordingly, in order for 35 percent of the targets to have been identified, approximately 35 percent of all of the pictures, including non-targets, must have been identified. However, Intraub's data are not unproblematic. First, there is a large difference between the detection rate for targets in positively (71 percent) and negatively (35 percent) phrased groups. Thus, it could be that set is playing some role in the positive group. Second, it could be argued that in the negatively phrased condition, subjects created an expectancy for the non-targets, along with a response rule to choose the stimulus that didn't meet that expectancy. Since all of the non-targets would meet such an expectancy (and would probably share some set of common features), such a strategy, though undoubtedly less efficient than responding positively to an expectation, would work. Finally, Intraub only used pictures of single objects, while Potter used pictures of both single objects and several objects. As was pointed out above, different processes may well be involved in the
identification of objects versus scenes, and for the purposes of the current discussion, it is the speedy identification of scenes that is of importance.

Finally, data provided by Potter (1976) can be taken as evidence in favor of an expectancy explanation for the speed at which her subjects could identify the target pictures. As noted above, it has previously been shown that one consequence of expectancy is to decrease the identification accuracy of those items which are not expected (Tulving & Gold, 1963). If expectancy is not the reason for the high rates of detection in the Potter paradigm, then recognition for non-targets should be equally good whether the sequence of pictures is viewed with or without a target search task. On the other hand, the expectancy hypothesis predicts that recognition memory for non-targets will be worse when the sequence is viewed with the target search task than without. Consonant with the expectancy hypothesis, recognition memory for non-targets was significantly lower when the subject was watching for a particular target than when not.

In summary, the data often cited as evidence supporting the view that the gist of a scene can be identified quickly is not as conclusive as it is sometimes portrayed. Admittedly, many of the alternative explanations given above are shamefully post hoc. Still,
while it does appear that a target picture can be detected quite rapidly (on the order of 100 msec), it is questionable whether this high rate of detection could occur without prior expectation. It is also questionable whether the non-target pictures (those for which there is no expectancy) are being processed to a semantic level at these short exposure durations. Finally, these experiments for the most part involve one or very few objects, not scenes, and therefore say nothing about how quickly complex scenes can be processed to a conceptual level. Therefore, it seems premature to conclude that the meaning of a scene for which there is no prior expectation can be accessed at a speed which would be useful for subsequent object identification.

Evidence for Schema Theory

The results most often cited (e.g. by Antes, 1977; Carr & Bacharach, 1976; Friedman, 1979; Loftus & Mackworth, 1978) as supportive of the claim that schema activation facilitates object identification are those of Biederman and his colleagues. In the paradigm experiment, Biederman (1972) presented photographed scenes and had subjects identify which object from a response set of four objects occupied a given cued position in the scene. The main variable of interest was whether the
scene was coherent or jumbled. Jumbled scenes were formed by dividing the coherent scenes into six equal sections and rearranging five of these (the sixth section, which contained the target object, was not moved). Other manipulations included stimulus duration, presentation order of the position cue and the scene, and presentation order of the response set and the scene. The results were that the scene coherency and the presentation order manipulations affected object identification. Identification was more accurate (a) with coherent scenes, (b) with the location cue presented prior to the scene, and (c) with the response set presented prior to the scene. Stimulus duration was not a significant factor, and there were no interactions. Based on these results, Biederman concluded, "It is most likely that jumbling affected an early, but not peripheral, stage involved in the perceptual recognition of the cued object," and wondered, "Is the functional unit an individual object, or does an observer have access to more global units or schema?" (Biederman, 1972, page 79). Based on further research employing the jumbling manipulation (Biederman, Glass, & Stacy, 1973), Biederman opted for the primacy of schemas: "an initial holistic categorization of the stimulus (determined by gross feature tests and context) biases the subsequent testing, weighting, and combination of detailed features..."
According to Biederman (1981) a schema for a scene can be activated by "global features" of the scene. These global features are what the jumbling manipulation is assumed to disrupt. The flow of information in this type of model is depicted in Figure 1.

The concept of global features as posited by Biederman cannot be restated as suggesting simply that spatial relationships affect object identification. Instead, global features are thought to be higher order visual features which allow access to schema representations in memory before objects and spatial relationships have been (fully) identified. The jumbling manipulation is thought to disrupt the global features in the scene, making schema activation slow or impossible. This analysis of course assumes the reality of such entities as global features. However, while Biederman (1981) has attempted to produce examples of global features, they have yet to be adequately specified. Further, it is difficult to see how such features could be specified, given the possible visual diversity of even the simplest of scenes. For example, the global features of a scene are taken to arise from the objects in the scene. Therefore, the global features incorporate information about both general shape and spatial position of the objects which comprise the...
Figure 1

The flow of information in a schema model.
scene. In order for these features to be useful, they would have to be fairly invariant (as is also true of any feature based theory of object recognition). Given the problem of finding invariant features in objects, the problem of invariant global features for scenes would seem even more difficult. For a scene such as a living room, the number, shape, and spatial arrangement of the objects are each extremely variable, and any of these factors would change the global features of the scene. It seems incumbent upon schema theorists to show how global features could be specified under these types of variations.

In the following discussion, an attempt will be made to show that the conclusions drawn from the experiments by Biederman and his colleagues may be unjustified. Several general problems with the basic methodology will be discussed.

Consider first the manner in which Biederman’s jumbling manipulation is performed. A photograph is cut into six equal pieces, and all but one of these pieces are then moved to a new location. The most striking thing about these jumbled scenes is the disruption of meaningful contours and the addition of new, meaningless contours. If contours play the central role in vision that many theorists believe (e.g. Hochberg, 1978; Julesz, 1971; Marr, 1982) then this manipulation may be
equivalent to a severe degradation of the stimulus through the introduction of meaningless visual noise. Under such circumstances, processes operative in object identification would be expected to suffer even under conditions in which the subject knew where to look and what to look for (as found in Biederman, Rabinowitz, Glass, & Stacy, 1974). Supportive of this claim is the fact that the jumbling manipulation interacted with stimulus duration in an object identification task (Biederman et al., 1974). While a schema theorist might like to conclude that an object in a coherent scene is identified faster due to the aid of a schema in the non-jumbled condition, an alternative explanation is that object identification takes longer given a degraded stimulus. (This issue could be resolved by testing object identification in coherent versus jumbled scenes in which the test object does not belong. The schema theory predicts that a coherent scene which does not activate a useful schema (i.e. a scene in which the object does not belong) will not facilitate object identification over the same jumbled scene, while the stimulus degradation position predicts that the jumbling manipulation will be about equally disruptive regardless of whether the object fits in the scene or not.)

A second problem with the jumbling manipulation is that the addition of meaningless contours and the
disruption of meaningful contours may be especially troublesome in the visual periphery where incoming information is not optimal even under the best viewing conditions. In experiments where subjects must scan a scene in search of a particular object (e.g. Biederman et al., 1973), eye movements play a significant role. Under such circumstances extrafoveal information is used to help determine where to send the eye next (Antes, 1974; Mackworth & Morandi, 1967). In addition, useful visual information gathered from the periphery can be integrated with foveal information when that area is fixated in order to facilitate identification (Pollatsek, Rayner, & Collins, 1984). The disruption of contours may adversely affect any of these processes.

The jumbling manipulation may further interact with other variables which affect the number of object identifications or eye fixations needed to carry out the task, as in Biederman et al. (1973), where the design of the experiment allowed a quick exit from an object search if the object was unlikely in the scene but required an exhaustive search if the object was likely. The results are easily explained by the extra processing time required in the jumbled condition for each additional object identification or eye fixation necessary. In fact, Biederman et al. (1973) were aware that the schema explanation was not necessary in order to account for
their data, since the presentation time was long enough to allow several eye fixations. However, even under conditions where eye-movements are not possible, such as in Biederman et al. (1974), a schema explanation is not necessary in order to explain interactions between the jumbling manipulation and other variables which affect the time needed to process the scene, since extra processing time is likely to be required in order to identify extrafoveal objects in the jumbled scenes. Again, an interaction between jumbling and a variable which affects the amount of time spent on extrafoveal processing (e.g. the label similarity manipulation in Biederman et al., 1974), can be explained (and would be expected) given that each additional "unit" of extrafoveal object sampling would require more time in the jumbled scene condition when compared with the normal scene condition. Given this analysis it does not seem necessary to postulate that the jumbling manipulation affects a schema activation process. Instead, the disruption of natural contours and concurrent stimulus degradation seem sufficient to account for the reported results.

In addition, a third problem with the jumbling manipulation is that it destroys the spatial structure of the scene. While on the face of it this may sound similar to stating that jumbling destroys "global features,"
there is an important distinction to be made. In the schema model, global features serve to suggest a schema, which then acts in a top-down manner to facilitate object identification (see Figure 1). An alternative to this account is to suggest that object identification and spatial structure identification are two separate processes which operate in parallel and without communication (i.e. they are informationally encapsulated with respect to one another; see Fodor, 1983; Marr, 1982; Ullman, 1985). If jumbling were to make the construction of a spatial representation of the scene more difficult (without necessarily affecting object identification), then tasks which require identification of the spatial relationships between objects (what was where?) would be more susceptible to performance inhibition from this manipulation. In the experiments conducted by Biederman and his colleagues requiring identification of an object at a cued position (Biederman, 1972; Biederman et al., 1974), the manner in which the subjects made their responses was to point to one of four possible choice objects. It is important to note that the three distractor objects were all in the scene. Therefore, this task requires not only object identification, but also a representation of object location. Thus, in these experiments it is not at all clear whether it is the object identification or spatial
structure process (or both) that is being affected by the jumbling manipulation.

In a later series of experiments, Biederman has attempted to show that "semantic" relations between objects are accessed at least as quickly as "physical" relations. In these experiments (summarized in Biederman, 1981), semantic relations (e.g. the probability of an object occurring, being in a particular position, and being a particular size) and physical relations (e.g. support for an object which should rest on something, and interposition or occlusion of an object when it occurs behind another) were manipulated in a scene. For example, in a living room scene, a floating couch would violate the physical relation of support, while an upside-down couch would violate the semantic relation of position. The effects of manipulating these relations were examined in object detection and violation detection tasks. Based on the findings that both types of relation violations equally affected performance in these tasks, and further that semantic relation violations added to the disruption in task performance caused by physical relation violations, Biederman concluded that "semantic relations are accessed at least as rapidly as relations reflecting the pervasive physical constraints of interposition and support that are not dependent on meaning..." (Biederman, 1981, page 253). Since
semantic violations affected object identification at least as much as physical violations did, the implication is that semantic relations are computed simultaneously with a physical parsing of the scene and before object identification.

There are many problems with these experiments, but two warrant special mention. First, much hinges on the distinction between support as a physical relation versus position as a semantic relation, yet this distinction seems neither empirically nor theoretically justifiable. On the one hand, if the human visual system processes support as a physical relation (at the level of the physical parser, e.g. Winston, 1970), then the same disruption that a floating couch causes should also be found with birds, balloons, airplanes, ceiling lights, and pictures hanging on a wall, none of which have visually obvious means of support. Biederman has provided no evidence for the position that support violations invariably cause disruptions in scene processing, and it is doubtful that those who believe in an early physical parsing level for human vision, such as Winston, would make this claim. On the other hand, if support is determined by a later stage of analysis which bases its decision in part upon the identity of an object (i.e. the semantics of the object), then the distinction between support as physically specified and position (or
probability or size) as semantically specified seems meaningless. Just as the detection of a position violation depends in part on the semantics of the scene, so too does the detection of support violations. Therefore, the finding that support violations affect the same stages of processing as violations of position or size or probability shows not that semantic representations of a scene are available as quickly as a physical parse of the scene, but instead shows only that one type of semantic relation is available at about the same time as others.

In addition to support, Biederman also included interposition as a physical relation. Interposition occurs when a solid object occludes the contours of an object behind it. Again, in these experiments it was found that the semantic violations were equally as disruptive on object identification as was an interposition violation (an interposition violation was defined as the ability to see the contours of another object through what should have been an opaque object, for example, being able to see a man's leg through his briefcase). Also, it was found that semantic violations were more easily and more quickly detected than was an interposition violation, and that semantic violations added to the disruption caused by an interposition violation. And again, based on these results, the
conclusion was that semantic relations must be accessed at least as quickly as physical relations.

One could, of course, wonder whether interposition is any less semantic than support--after all, glass is a fairly common substance. Still, even if it is taken as fact that interposition is determined at a strictly physical level, questions remain. First, it is worth noting that in general, across all of the experiments discussed in Biederman (1981), interposition violations have little if any effect at all. If one examines the examples of the line drawings used in these experiments, one possible reason for this becomes apparent: Interposition violations are very difficult to see in line drawings.

However, even ignoring the above criticisms, a possibly fatal methodological problem presents itself: In these experiments, a brief view of the scene is followed by a pattern mask, and the effects of semantic versus physical violations are then examined. Yet, as Marcel (1983) has shown, pattern masks differentially affect various representational levels. In particular, pattern masks seem to affect the availability of physical representations of visual input while leaving semantic representations relatively unaffected. Given this possibility, the use of a pattern mask in experiments attempting to contrast the effects of physical and
semantic representations seems ill advised, and one is left wondering what would have happened in these experiments if a pattern mask had not been employed.

A final problem with both the violation and jumbling paradigms is the nature of the tasks used to explore object identification. Object identification is one stage in a series of processing stages required for the construction and retention in memory of the representation of a scene. In order to isolate the object identification stage from later stages in the processing sequence, it is necessary to choose a measure of object identification carefully. Most preferably, one would use an on-line measure of performance, where an on-line measure can be defined as one which taps a representation as it is being constructed. With regard to object identification, such a measure should be unaffected by processes and representations which occur after the object identification stage. None of the experiments reviewed in this section employed an on-line measure, and it is not clear that post-perceptual information integration and response bias effects have been adequately controlled. In the next section, the results from experiments using a different measure, eye movement patterns recorded during scene processing, are examined.
Eye Movements and Picture Viewing

One particularly useful method for studying the effects of a scene context on object identification, and for studying scene processing in general, is to record the eye movements of the viewer. This technique allows the experimenter to know exactly where and for how long an observer is viewing a particular section of the scene.

During picture viewing, eye movements can be characterized as consisting of two components, fixations and saccades. Fixations are short periods of time averaging between 250-350 msec (Rayner, 1978; Yarbus, 1967) during picture viewing when the eye is relatively stationary and visual information is taken in. Saccades are quick jumps of the eye which occur between fixations and which bring new information into the fovea. Neither fixations nor saccades are homogeneous during picture viewing; fixation durations, fixation densities, and saccadic extent are all quite variable, even for one observer viewing one picture for a short amount of time. In reading, fixation duration and saccadic extent seem to reflect different underlying cognitive processes (e.g., Rayner & McConkie, 1976), and this may also be true in picture perception, though it has yet to be shown empirically.
Based on the results from studies recording eye movements during reading (see Rayner, 1978), it seems that eye movement patterns during picture perception provide an excellent "online" measure of the time course of underlying cognitive processes. It has further been assumed that the variability in fixation durations and saccadic extents provides a good indication of processing difficulty. For example, First Fixation Duration, the duration of the first fixation on an object or section of a scene, has been taken as a reflection of the time needed to encode that object or section (Friedman, 1979; Loftus, 1972). (In contrast, Gaze Duration is the total fixation time across saccades spent fixating an object or section of the scene, and is taken to reflect both encoding and post-encoding processes.) While it may be that factors other than encoding difficulty affect first fixation duration, it seems a reasonable assumption that first fixation duration reflects at least encoding time. Since the primary interest in this paper is object identification or encoding, the preferred variable will be first fixation duration.

Context Effects on First Fixation Duration

Under the assumption that first fixation duration is a good reflection of encoding time, the hypothesis that schemas mediate contextual effects on object
identification predicts that the first fixation duration on an object which is an explicit argument in an activated schema will be shorter than the first fixation duration on an object which is optional to the activated schema, which in turn should be shorter than the first fixation duration on an object which does not fit the schema at all.

Loftus and Mackworth (1978) conducted a study in which they constructed 78 groups of four pictures, the four pictures consisting of a pair of normal scenes and a pair constructed by switching one object from each of the scenes into the other (thus a cow in a farm scene would be switched with an octopus in an underwater scene) to create scenes with one "informative" or low probability object in each. Subjects were then shown one of the four scenes from each group for four seconds each and told to examine the picture "as if" for a later recognition test. Consistent with the above prediction, low probability objects were fixated longer at the same location in the same background as high probability objects, both on the first fixation to the object and on the subsequent few fixations. Since the distance of the fixation just prior to the fixation on the object of interest was about the same for both conditions (an average of 6.5 to 8 degrees of visual angle), the difference in first fixation duration was not accounted
for by a closer parafoveal preview in one condition than another. It was therefore concluded that encoding time for objects which fit a schema are faster than for objects which do not.

There are, however, several possible problems here. First, even though the average distance of the last fixation before the fixation on the target object was the same regardless of condition, it is still possible that more visual information was processed parafoveally when the object fit the context (see Balota, Pollatsek, & Rayner, 1985, for evidence of this in reading). That is, low probability objects may be fixated longer on the first fixation because they are less thoroughly processed in the parafovea. Of course, this leads one to ask why more information can be extracted from a probable object in the parafovea, and the schema theory may equally well take this type of data as supportive. A second problem is that while Loftus and Mackworth show that objects which do not fit in a scene at all are fixated longer than objects which are highly likely to be in the scene, they do not examine the second part of the prediction, which says that the first fixation duration on an object which is an explicit argument in an activated schema will be shorter than the first fixation duration on an object which is an optional argument in that schema. Perhaps longer first fixations on objects
which don't fit in a scene at all (e.g., octopus in a farm scene) reflect not longer encoding time, but a "double-take" or rechecking of the input to make sure that the original encoding was correct.

In a study conducted by Friedman (1979), one group of subjects rated the likelihood that each object from a list of objects would be found in a particular scene. Complex line drawings were then constructed which contained high, medium, and low probability objects. Unlike Loftus and Mackworth (1978), all of the objects were at least reasonable in these scenes (e.g., a fireplace in a kitchen would be a low probability object). A second group of subjects then viewed each of six pictures for 30 seconds. The pictures were precued by the general topic (e.g., "kitchen") and the subjects were told that they "would later have to be able to distinguish between the original pictures and new pictures in which, for example, only a small detail on one object would be different." Again, consistent with the schema hypothesis, first fixation durations to high probability objects were significantly shorter than first fixation durations to either medium or low probability objects. Further, in support of the notion that first fixation duration reflects perceptual factors such as encoding better than later fixations, rated likelihood accounted for far more of the variation in fixation
duration on the first fixation (between 27.7% and 52.4% depending on the scene) compared to the second (14.4% to 38.6%) or later (5.6% to 24.1%) fixations, and though there was an association between rated likelihood and total fixation duration or "gaze" (324 msec, 379 msec, and 562 msec to high, medium, and low probability objects respectively, according to Friedman and Liebelt, 1981), this association virtually disappeared when the variance attributable to first fixations was partialled out.

In criticism of this study, it could be argued that 30 seconds is an unusually long time to view continuously a line drawing of a scene, and that because subjects knew they would have this amount of time for each picture, they may have adopted an unusual viewing strategy. Second, Friedman precued subjects on the topic of each scene, so the ability to access a schema quickly and use it in object identification wasn't tested. On the other hand, it may be that precuing has little effect on schema activation (Biederman, Teitelbaum, & Mezzanotte, 1983), and even if it does, it could be argued that precuing is more ecologically valid since one normally has some idea of the general nature of the next visual event. Third, since there was no significant difference between first fixation duration to medium and low probability objects, the prediction of the schema theory was not fully supported. Still, null results are not very
convincing either way, and two of the three comparisons were in the predicted direction. In general, these results seem to support Loftus and Mackworth (1978) in the conclusion that objects which are more likely to appear in a particular scene (and so be arguments in a particular schema) are fixated for less time on the first fixation than objects that are less likely to appear.

There is a further problem common to both of the above studies. According to the schema theory, the encoding of objects which are arguments in a particular schema should be facilitated when that schema has been accessed. Therefore, the prediction of the schema theory is not only that the first fixation duration to a highly likely object will be shorter than that to a less likely object, but also that first fixation duration to an object appearing in a scene in which it is highly probable will be shorter than when it appears in a neutral-context baseline. On the other hand, it is not clear whether encoding of an object which does not fit with the schema should be inhibited relative to a neutral baseline; the prediction would depend upon the process by which schemas affect encoding. If target objects which do not fit in a scene require active hypothesis testing mechanisms, as proposed by Friedman (1979), then an inappropriate schema activated by an inappropriate scene would be expected to cause inhibition in identifying that
object in comparison to a neutral context, since the inappropriate schema would suggest inappropriate hypotheses to be tested. If, instead, schemas affect object encoding through automatic processes, then no inhibition would be expected. While both Loftus and Mackworth (1978) and Friedman (1979) have shown that expected objects are fixated for less time relative to low probability objects, it could be argued that this difference is due to factors slowing down processing of low probability objects rather than speeding processing of high probability objects. Since the schema theory specifically predicts the latter, this distinction is important.

A recent study by Antes and Penland (1981) attempted to provide a neutral baseline condition in order to determine whether providing the context of a scene was in fact facilitating encoding of expected objects as predicted by the schema theory, was inhibiting encoding of unexpected objects, or both. They constructed 23 complex line drawings of scenes, within each of which four objects were designated as targets. One pair of targets in each picture was within 5 degrees of the center of the picture while the other pair was beyond 5 degrees, and each pair contained one expected and one unexpected object as determined by subjective probability ratings. Aside from these 23 scenes which were designated
the high context condition, 23 low context pictures were created by copying the four target objects plus two other non-overlapping objects from the scene, preserving spatial positions, but with no other objects or background. Ten pictures from each context condition were shown in a random order for four seconds each, and subjects were told "to view the picture in preparation for an object recognition test after each picture." This test involved choosing the object which had appeared in the picture from an array of four objects.

Consistent with the prediction of the schema theory, there was a significant interaction between context and expectedness, such that mean first fixation duration on expected objects was shorter in the high context condition than in the low context condition, while there was no difference for unexpected objects across the two conditions. Taking the low context condition as a baseline, it therefore appears as though encoding of an object which is expected in a given scene is facilitated relative to a base rate of encoding while an object which is not expected is neither facilitated nor inhibited. Before accepting these results as definitive, however, several aspects of the study should be noted. First, as in the Loftus and Mackworth (1978) study, many of the unexpected objects were not only unexpected, but simply did not fit in the high context
scenes at all. Second, as Antes and Penland point out, contrary to what might have been predicted given the schema theory, the mean first fixation duration on expected objects was not significantly shorter from that on unexpected objects in the high context condition. Still, this null finding should not be too worrisome since the data were in the right direction, and since it appeared as though the expected objects were generally more difficult to encode than the unexpected objects, as indicated by the finding that first fixation durations to the expected objects were longer in the low context condition also.

A third and more fundamental problem with the Antes and Penland study is that there were more objects in the high context than low context condition. Because of this, a direct comparison across the two conditions becomes difficult. For example, since there were more objects in the high context condition, the first fixation on a target object in the high context condition would presumably be later on the average than in the low context condition. This means that there would have been more parafoveal processing of the target objects in the high context condition The fact that there was a significant main effect of context such that first fixation duration for both expected and unexpected objects was shorter in the high context condition than in
the low context condition lends support to this possibility. If highly constrained perceptual stimuli benefit more from parafoveal processing than do less constrained (as suggested by the results obtained in reading by Balota, Pollatsek, and Rayner, 1985), then differences in first fixation duration to target objects may reflect different levels of parafoveal priming rather than directly reflecting the effect of context on absolute encoding time. As pointed out above, a schema theory advocate could argue that this explanation simply moves the problem back a step and leads one to ask how context allows extraction of more information from a highly constrained object in the parafovea than from a less constrained object. And schema theory may posit a similar explanation for this type of effect as for differential first fixation duration, without compromising the position that a schema facilitates encoding. However, another consequence of the difference in the number of objects in the two conditions is that the estimate of encoding time in the baseline condition may have been inflated. This could occur for two reasons: (1) subjects may have adopted a strategy of spending more time on each object in the baseline condition because they knew that they would have the same total amount of viewing time but fewer objects to look at, and (2) fewer objects in the parafovea may
"draw" the eye less powerfully away from the foveal stimulus. In either case, encoding time without context would be overestimated. If the baseline were overestimated, then the pattern of facilitation for high probability objects but not for low probability objects would be an artifact. Until the two conditions are more tightly controlled, it seems ill-advised to conclude one way or the other whether context is facilitative for high probability objects, inhibitive for low probability objects, or both.

Overall, the above three studies taken together indicate only that first fixation durations to high probability objects are on the average shorter than those to low probability objects. While this result is consistent with schema theory, it is probably also consistent with a multitude of other theories as well. In fact, any memory structure/process combination which could generate expectations about which objects should be encountered next (given a certain scene or set of objects as context) would be equally predictive and explanatory of the data. Since the few obvious predictions distinguishing schema theory have yet to be adequately tested, it is not at all clear that one need specifically posit schemata as the type of memory structure responsible for the context effect, even if one feels that a top-down influence from some activated memory
structure is the cause of these effects. (It should be obvious that data which seem to indicate that memory for scenes involves schema use, e.g., Antes, 1977; Friedman, 1979; Mandler and Johnson, 1976, do not bear on the issue of object identification in scenes, since memory for and perception of scenes may entail the use of totally different structures and processes.)

Further, there may be good reason not to use the theoretical construct of a schema at all, since it is extremely broad and undiagnostic. Friedman (1979) gives the most precise account of the manner in which schemata are thought to affect perceptual processing, and even here there are few predictions which can be made without argument. For example, Antes and Penland (1981) suggest that expected objects demand more attention in the parafovea (i.e., should be fixated sooner), while Loftus and Mackworth (1978) suggest that unexpected objects do. Both pairs of researchers base these predictions on the schema theory. Also, the exact mechanism by which a schema produces contextual facilitation has yet to be specified. So far, only the distinction between top-down and bottom-up processing has been suggested to account for object identification in and out of context. Without a more explicit model of how top-down knowledge can affect object identification, the schema theory will remain extremely difficult to
General Criticisms and an Alternative Explanation

Aside from the specific problems for schema theory raised in the above sections, there are also some general considerations that should be taken into account in forming a theory of context effects and visual object identification. Perhaps chief among these is the degree to which the theory allows context to influence object identification. For many years, the common wisdom held in computer vision was that all knowledge sources would have to be consulted in a highly interactive way in order to identify an object. This type of view was accepted into cognitive psychology (e.g. Neisser, 1967) and is one of the traditions from which schema theory in visual cognition derives. However, models of this type have severe problems, one of which is that systems which rely heavily on top-down processing tend to see what they expect to see rather than what is actually in the environment. And as Fodor aptly says, "a condition for the reliability of perception, at least for a fallible organism, is that it generally sees what's there, not what it wants or expects to be there. Organisms that don't do so become deceased" (Fodor, 1983).

Since schema theories in psychology are generally
not specified in great detail, it is often difficult to
tell the exact manner in which the top-down influence
from the schema is supposed to affect the identification
process, and therefore to what degree the top-down
information actually influences the content of the
resultant perceptual descriptions. Certainly many
psychologists who believe in schema theories of vision
also believe that misapplication of a schema will lead to
misidentification of objects (e.g. Treisman & Gelade,
1980). However, one of the most remarkable aspects of
human vision is the rapidity with which objects can be
recognized, even in the absence of expectation
(e.g. Biederman, 1985). In fact, when misidentification
does occur, it is usually quite striking, for example,
when a cardboard box on the road is taken as an animal
under the degrading visual conditions of a dark
night. The fact that misidentification is striking when
it occurs suggests that it doesn’t occur very often. The
question that needs to be asked is whether it is
reasonable to posit that the human visual system is as
easily fooled as schema theories seem to imply.

A second question which schema theory has yet to
address is how the knowledge contained in a schema
actually influences the recognition process. Some
possibilities are that a schema alters the order in which
memory representations are matched against the input,
causes a search for particular features or parts of objects in particular places, lowers the goodness of fit threshold for expected objects, generates and fits particular expected templates, or fills in expected parts of objects (Pinker, 1984). Friedman (1979) has been most explicit in her description of how a schema facilitates object recognition (see discussion above), but it is not at all clear that other researchers accept her view. Certainly the predicted effects of schema activation are going to be different depending on the manner in which the knowledge contained in a schema is used. It is an issue whether schema theory can be taken as a theory at all until it is specified to a level that allows testable predictions and consequences to be formulated.

A third type of problem for any theory of perception (including schema theory) which allows general semantic knowledge to influence perceptual processes, is what has come to be known as the "frame problem" in artificial intelligence (Minsky, 1975). In brief, the problem is one of determining which aspects of all available world knowledge are relevant to a particular situation. For example, Biederman, Mezzanotte, and Rabinowitz (1982) specify that position is a semantic relation of objects to their visual contexts that is contained in the schema used in object identification. However, how much position
knowledge is contained in the schema? Presumably, the
schema for a living room would dictate that chairs should
be on the floor. Would this then predict that a chair
placed on a couch (say, so that the floor could be
washed) would be difficult to identify? Or is the
knowledge that chairs are moved when floors are washed
also contained in the schema? If someone picked up a
chair to move it, would this make it harder to recognize? Must the schema contain the fact that people
can lift chairs? The issue, then, for schema theory is to
provide an account of how and where a boundary is
drawn around the relevant information given a particular
context. So far, theorists working in computer vision
have found this problem extremely difficult, and
consequently much of their work has returned to a more
stimulus driven approach (Marr, 1982).

A final issue relates to the type of theorizing that
one would want in the study of vision. If the choice is
made to deal with all difficult problems in vision by
postulating top-down processing, then processes which in
fact are computed in a more bottom-up fashion may be
totally missed. Further, an emphasis on top-down
processing will cause an under-estimation of the
information that is contained in the light array reaching
the eye (e.g. Gibson, 1966; Marr, 1982). Another way of
saying this is that we are more likely to discover
structure by postulating its existence than by assuming that it doesn't exist (as argued by some researchers working in language processing, e.g. Forster, 1979; Frazier, Clifton, & Randall, 1983). If, after a careful search, it turns out that certain problems are impossible to solve without postulating top-down influences, then at that time models should include them. But if models start out assuming top-down processing, then the representations which are computed in a bottom-up fashion and the processes that compute them may never be found.

If the assumption that top-down influences are ubiquitous in object recognition is rejected, then how are the demonstrated effects of context on object recognition to be explained? One possibility is that all of the effects are actually due to post-perceptual factors, such as response biases induced by the tasks used to study recognition, and semantic integration of the recognized objects into a higher level representation of the entire scene. The former explanation can probably be ruled out, at least for those studies which employed eye movement monitoring, since in those experiments there were no overt responses required. However, the latter explanation is more difficult to rule out; if an object doesn't fit into its scene, then presumably it is going to be more difficult to integrate that object into a semantic representation of the scene as a whole. To the
extent that the techniques used to study recognition employ measures susceptible to post-perceptual processing such as integration, it is going to be difficult to tell whether recognition itself can be influenced by context. There is no guarantee that either the detection paradigm used by Biederman or the fixation duration measure are influenced by recognition factors alone.

An alternative explanation for the effects of context on object identification derives from the literature on object priming. This work has its foundations in the classic lexical priming effect demonstrated by Meyer, Schvaneveldt, and Ruddy (1975) in which it was found that recognition of a word such as "nurse" was facilitated when that word was preceded by a related word such as "doctor" rather than when preceded by another, unrelated word. The explanation for this effect is taken to be automatic, passive spreading activation between nodes in lexical or conceptual memory (Collins and Loftus, 1975).

Using a paradigm similar to that used in examining lexical priming, similar effects have been found with pictures of objects as stimuli (e.g. Carr, McCauley, Sperber, & Parmelee, 1982; Kroll & Potter, 1984, McCauley, Parmelee, Sperber, & Carr, 1980). Again, the explanation for this effect is generally agreed to be passive spreading activation between nodes in semantic
memory. Therefore, an alternative explanation for context effects seems to be available. This explanation will be denoted as "intralevel priming".

Intralevel Priming

According to this view, context can affect object identification by virtue of a mechanism in which already identified objects prime representations for related and as yet unidentified objects. This priming could take place intralexically (through representations of the object names) or at the level of object representations. In either case, a higher level representation of the entire scene is not invoked. The mechanism for this priming can be thought of as similar to that proposed in the word recognition literature, such as spreading activation (Collins & Loftus, 1975). Further, these processes are hypothesized to be automatic (Posner & Snyder, 1975), as opposed to requiring attentional resources. In its extreme form, this type of explanation is equivalent to the proposal that object identification is performed by a processing module which is "informationally encapsulated" (Fodor, 1983). That is, the object identification processor is deaf to representations which are formed on the basis of the module's output. Therefore, higher level processors which integrate the output provided by the module cannot
provide input to that module nor affect the module's subsequent processing of new input. Said another way, a modular system would not allow the integrated representation of a scene to affect the processing of an as yet unidentified object in that scene.

The types of information which a module can use are representations which must be computed prior to the operation of that module and which serve as its input, for example featural information in the case of object identification, and information which is necessarily contained within the module, for example the object representations themselves and the structure which organizes them and makes them retrievable.

According to the intralevel priming view of object identification, context effects occur as a result of the organization of the representations within the module. Simple associative or semantic links among the object representations allow priming among objects without top-down processing. While this seems to make the object recognition module relatively "dumb", the tradeoff is that the module gains the benefit of speed: All sources of information need not be considered in computing an output. Evidence for the quick access of meaning from individual objects has been provided by the Potter experiments discussed above, as well as by Biederman (1985). Whether foveal objects would similarly
prime subsequently fixated objects is the question addressed by these Experiments.

If it can be shown that a single object fixated on fixation $n$ can prime the identification of a related object on fixation $n+1$, then the beginning of an explanation for the effects of context on object identification in scene processing can be proposed which does not involve the use of schemas. The explanation would be that in conditions where an object fits a scene context, the likelihood will be high that a semantically related object will have been fixated immediately before the target object was fixated. Spreading activation from the node of the object identified on fixation $n$ would then raise the activation level of all related object nodes. When, on fixation $n+1$, a related object was fixated, recognition of that object would be facilitated due to the higher level of activation of its representation node. On the other hand, if an object is placed in a scene in which it does not fit, the probability will be relatively high that an unrelated object will have been fixated immediately before. Therefore, the node representing that object will be at a resting level of activation and will require more extensive processing in order to be recognized. This simple mechanism will therefore account for the longer fixations needed to recognize objects which do not fit
the scene context. (In experiments which do not allow eye movements, such as Biederman et al., 1974, the assumption would be that in a proper scene context, the foveal object will be more likely to be related to and to prime the parafoveal identification of the target object.) This explanation makes the assumption that objects which go together in scenes tend to be more semantically related to each other than objects which do not form a scene. A moment's reflection will show that this assumption is not unreasonable.

The two experiments presented in this thesis are an attempt to explore the effects of context on the recognition of objects using a paradigm which has been shown in the word priming and object priming literature to be sensitive to perceptual and relatively insensitive to post-perceptual factors (Seidenberg, Waters, Sanders, & Langer, 1984). This paradigm uses naming latency as an indication of recognition time. In addition, these experiments attempt to provide a direct test of the intralevel priming model of context effects by using only a single foveal object as context.
It has been repeatedly demonstrated that identification of a pictured object presented foveally is facilitated when that object is preceded by a related object also presented foveally, compared with when an unrelated object precedes the target object (e.g. Carr, McCauley, Sperber, & Parmelee, 1982; Kroll & Potter, 1984; McCauley, Parmelee, Sperber, & Carr, 1980). The results of these experiments mirror the effects found when both associated and semantically related words are used as stimuli (e.g. Meyer, Schvaneveldt, & Ruddy, 1975; Fischler, 1977; Henderson & Hansen, in progress), and are generally interpreted as an automatic process (Posner & Snyder, 1975) within a spreading-activation framework (Collins & Loftus, 1975).

In addition, a second set of experiments, conducted by Pollatsek, Rayner, & Collins (1984) has shown that information gathered from an extrafoveal object during one fixation can facilitate the recognition of that object when it is foveally fixated (replicated by Pollatsek, Rayner, & Henderson, in progress). The Pollatsek et al. studies demonstrate that information
from an extrafoveal object can be integrated across saccades. These results also mirror those found in studies with words (Balota & Rayner, 1983; McConkie & Zola, 1979; Rayner, McConkie & Zola, 1980.)

The above studies are important because they indicate that the visual context in which an object occurs influences the identification of that object, and yet the explanation for these effects does not invoke schema-type processes. However, a problem with both of the above paradigms is that they are only approximations to the sequence of events which occurs during normal scene viewing. The foveal priming studies can be thought of as simulating eye movements, but unlike normal picture viewing, no eye movement is made and no extrafoveal preview information is available. The integration experiments are closer to normal viewing since eye movements are made to extrafoveal objects; however, in these studies, there is no meaningful object in the fovea before the eye moves.

The current experiment was an attempt to combine the above paradigms in order to determine whether an object fixated in the fovea would facilitate identification of a related extrafoveal object after a saccade brought the extrafoveal object into the fovea. This type of an effect would be expected if the intralevel priming explanation of facilitation found for object recognition in scenes is
correct. In order to examine this question, two stimuli were presented on the CRT simultaneously, one foveally and the other extrafoveally. On one third of the trials, the critical object pairs were related, one third contained unrelated objects, and one third involved a non-meaningful foveal "blob". The task was to execute an eye movement to the extrafoveal object and name it as quickly as possible. If priming is possible from foveal to extrafoveal objects, then naming times should be faster when the objects are related compared with when they are unrelated. (The schema theory makes no explicit predictions about what should happen here, although as currently formulated, there is no mechanism to account for such priming if it should occur.)

A concurrent purpose was to determine whether more information can be gathered from the extrafoveal object when there is a related object in the fovea. Such an effect of foveal context on parafoveal information extraction has been found with word targets, using both single word (Inhoff, 1982, Balota & Rayner, 1983) and sentence contexts (Balota, Pollatsek, & Rayner, 1985; McClelland & O'Regan, 1981). In order to examine the effect of having a parafoveal preview of the eventual target on identification once that target had been fixated, half of the trials allowed a parafoveal preview of the target, while on the other half of the trials no
parafoveal preview was given.

Finally, two other variables common to scene perception were manipulated: First, subjects participated in two blocks, one in which eye movements were from left to right, and the other in which eye movements were from right to left. Second, two visual eccentricities were used between the foveal and parafoveal stimuli: 5 degrees and 10 degrees. Several studies have shown differential effects of context on object identification depending on the distance of the to-be-identified object from the current fixation (Antes, 1974; Friedman, 1979; Parker, 1978). Also, the amount of parafoveal information extracted has been shown to depend on visual distance (Nelson & Loftus, 1980; Pollatsek et al., 1984).

The intralevel priming explanation of context effects in scene perception predicts that naming times in the related foveal prime condition will be faster than those in the unrelated foveal prime condition. A finding of this type would indicate that encoding a foveal object on fixation N can affect the encoding of another foveal object on fixation N+1. This could be interpreted as indicating: (1) Schema explanations are not necessary to explain context effects in scene viewing, since the same type of effect can be obtained with single object contexts (this would be an extremely strong conclusion to draw from these data-- however, the burden of proof might
then fall on schema theorists to show how a schema explanation adds to this explanation of context effects); or (2) A schema can be activated on the basis of a single object (this would require major modification of the schema model).

The Pollatsek et al. studies (1984; in progress) cited above have shown that a parafoveal preview of an object facilitates subsequent encoding of that object when it is fixated. Such an effect would also be expected in the present experiment. Of greater interest is how the effects of Foveal Prime and Parafoveal Preview may combine. An overadditive interaction between these factors (i.e. more priming when there is a preview) would imply that more information can be obtained from a parafoveal object when that object occurs in the context provided by a single related foveal object. If a single foveal object were able to both facilitate information extraction from a parafoveal object and facilitate encoding of that object once it was fixated, the need for postulating schema activation in order to account for context effects in scene viewing would be greatly reduced. On the other hand, if Foveal Prime and Parafoveal Preview were to show additivity with respect to naming time, additive factors logic would suggest that these factors affect different stages of processing (Sternberg, 1969).
Method

Subjects. Eight members of the University of Massachusetts subject pool participated in the experiment.

Materials. The stimuli were 60 line drawings of common objects which had been combined into 30 pairs of related objects, all easily identified and named (a complete list is given in Table 1; the line drawings were mostly taken from Snodgrass & Vanderwart, 1980, with a few exceptions). The same drawings were also randomly combined into 30 pairs of unrelated objects to serve in the unrelated foveal prime condition.

In addition, two control stimuli were employed: (1) A square, slightly larger than the objects, empty except for a small fixation cross in the center, was used as a parafoveal stimulus in the No Parafoveal Preview condition in order to give subjects a target to move their eyes to; and (2) A meaningless, roughly rectangular "blob" made up of irregularly drawn sides and filled with three irregularly drawn interior line segments, equated for the number of pixels contained by the average object drawing, was used as a non-meaningful foveal prime.

Subjects were asked to name each of the objects before the experiment. If necessary, the experimenter corrected the name employed, and the objects were
repeatedly presented until the experimenter was sure that the subject had the appropriate name for each object.

**Apparatus.** The stimuli were displayed on a Hewlett-Packard 1300A CRT with a P-31 phosphor. The CRT has the characteristic that removing a point results in a drop to 1% of maximum brightness in 0.25 msec. A black theater gel covered the CRT so that the display appeared clear and sharp to the subjects.

Eye movements were monitored via a Stanford Research Institute Dual Purkinje eyetracker. The eyetracker and CRT were interfaced with a Hewlett-Packard 2100 computer which controlled the experiment. The drawings were entered into the computer via a Summagraphics Bit-Pad. During the experiment, the computer kept a complete record of saccade latencies, accuracy, and naming latencies. The signal from the eyetracker was sampled every 1 msec by the computer and the position of the eye was determined every 4 msec. When the subject made an eye movement in the appropriate direction, the computer immediately replaced the parafoveal preview item with the parafoveal target object. The computer initiated the change when an eye movement of 0.5 degrees in the appropriate direction was detected and the change was completed within 5 msec. Since a saccade of 5 degrees (to the nearest target object) requires approximately 35 msec, the display change was always completed during the
saccade when vision was suppressed. Previous studies using this technique indicate that subjects do not see the change taking place.

The subject's eyes were 46 cm from the CRT and each object subtended approximately 2 degrees of visual angle horizontally and from 1 to 3 degrees vertically over the set of objects. Eye movements were monitored from the right eye, although viewing was binocular.

Procedure. Upon arriving for an experimental session, each subject was seated comfortably with his or her head resting on a chin and forehead rest to minimize any head movements. The room was dark except for the displays on the screen and a dim indirect light source. The calibration of the eye movement system then took place. After calibration, 32 practice trials were given followed by two blocks of 360 test trials. A trial consisted of the following events: First, a fixation display appeared (initiated by the experimenter), made up of five fixation crosses--one at the left-hand boundary of the screen, one at the center, one at the right-hand boundary of the screen, and one each approximately one degree toward the center of the screen from the boundary crosses. The subject was instructed to look at one of the fixation crosses (the inner-left cross in the left-to-right eye movement condition and the inner-right cross in the right-to-left eye movement condition), and
the experimenter checked to see whether the calibration was accurate. (A sixth cross moved with the computed eye position, indicating that the apparatus was calibrated if this sixth cross coincided with the appropriate fixation cross.) If the calibration was satisfactory, the experimenter warned the subject that the trial was to begin, and approximately 250 msec later the fixation crosses were replaced by a foveal stimulus (object or blob) and a parafoveal stimulus (object or box). The subject then moved his or her eyes to the parafoveal item. During the saccade, the parafoveal item was replaced by the parafoveal target object (as described above), and the subject named this target object as quickly as possible. The computer recorded the latency of the vocal response (timed from when the eye crossed the 0.5 degree threshold point). The experimenter recorded the accuracy of the response and/or whether there had been a loss of tracking accuracy on that trial.

The Eye Movement Direction factor was blocked, with the order of blocks counterbalanced across subjects. The experiment was completed in two sessions, one session for each block, generally run on consecutive days, with each session lasting 45 to 60 minutes.

Design

The experiment consisted of 720 trials per
subject. The 720 trials consisted of the factorial combination of 30 (parafoveal targets) X 3 (related versus unrelated versus non-meaningful foveal prime) X 2 (parafoveal preview versus no parafoveal preview) X 2 (5 degree versus 10 degree visual eccentricity) X 2 (left versus right eye movement direction). All factors were manipulated within subject. Eye Movement Direction was blocked, with order of blocks counterbalanced across subjects. Each foveal prime occurred with a related and unrelated parafoveal target an equal number of times in order to preclude conscious prediction strategies.

Table 1
Target objects and their related foveal primes.

<table>
<thead>
<tr>
<th>Foveal Prime</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>hand</td>
<td>foot</td>
</tr>
<tr>
<td>dog</td>
<td>cat</td>
</tr>
<tr>
<td>coat</td>
<td>hat</td>
</tr>
<tr>
<td>bee</td>
<td>flower</td>
</tr>
<tr>
<td>horse</td>
<td>cow</td>
</tr>
<tr>
<td>knife</td>
<td>gun</td>
</tr>
<tr>
<td>doctor</td>
<td>nurse</td>
</tr>
<tr>
<td>hammer</td>
<td>saw</td>
</tr>
<tr>
<td>truck</td>
<td>car</td>
</tr>
<tr>
<td>fridge</td>
<td>stove</td>
</tr>
<tr>
<td>lock</td>
<td>key</td>
</tr>
<tr>
<td>------</td>
<td>-------</td>
</tr>
<tr>
<td>leaf</td>
<td>tree</td>
</tr>
<tr>
<td>shirt</td>
<td>tie</td>
</tr>
<tr>
<td>sock</td>
<td>shoe</td>
</tr>
<tr>
<td>table</td>
<td>chair</td>
</tr>
<tr>
<td>fork</td>
<td>spoon</td>
</tr>
<tr>
<td>leg</td>
<td>arm</td>
</tr>
<tr>
<td>rabbit</td>
<td>squirrel</td>
</tr>
<tr>
<td>bat</td>
<td>ball</td>
</tr>
<tr>
<td>cheese</td>
<td>mouse</td>
</tr>
<tr>
<td>lightbulb</td>
<td>lamp</td>
</tr>
<tr>
<td>apple</td>
<td>pear</td>
</tr>
<tr>
<td>glass</td>
<td>cup</td>
</tr>
<tr>
<td>horn</td>
<td>drum</td>
</tr>
<tr>
<td>ashtray</td>
<td>pipe</td>
</tr>
<tr>
<td>comb</td>
<td>brush</td>
</tr>
<tr>
<td>wagon</td>
<td>sled</td>
</tr>
<tr>
<td>star</td>
<td>moon</td>
</tr>
<tr>
<td>bell</td>
<td>whistle</td>
</tr>
<tr>
<td>anchor</td>
<td>boat</td>
</tr>
</tbody>
</table>
Results and Discussion

The corrected mean naming latencies, collapsed over items, subjects, and direction of eye-movement, are presented in Table 2. Naming errors were very infrequent (they occurred on less than .01 of the trials) and were randomly distributed across conditions. The analyses reported in this section were conducted on corrected mean response times. These corrected times excluded all noise trials. Noise trials consisted of either (1) trials on which voice key failures, track losses, and naming errors occurred; (2) trials on which the saccade latency was either less than 150 msec or greater than 400 msec; and (3) trials on which naming latency was greater than 3 standard deviations from that subject’s mean latency for that particular block. The mean percentages of noise trials so defined for each condition are also shown in Table 2. The pattern of results for the corrected mean naming latencies did not differ from the pattern before correction.

A 3 (Foveal Prime) X 2 (Parafoveal Preview) X 2 (Eccentricity) X 2 (Eye Movement Direction) within-subject ANOVA was conducted on the mean naming latencies, treating subjects as a random effect.

Eye Movement Direction produced neither a main effect ($F < 1$) nor interacted with any other factor, and hence will be ignored in the remainder of this
The first question of interest is whether the two main variables, Parafoveal Preview and Foveal Prime, had an effect on naming latencies. Consistent with Pollatsek et al. (1984), there was a significant main effect of Parafoveal Preview \( F(1,7) = 357.4513, p < .00005 \). The mean naming latency was 647 msec with a preview and 723 msec without. In addition, the amount of benefit derived from a parafoveal preview was mediated by the distance of

Table 2

<table>
<thead>
<tr>
<th></th>
<th>No Preview</th>
<th></th>
<th>Preview</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Eccentricity</td>
<td>rel</td>
<td>unrel</td>
<td>blob</td>
<td>rel</td>
</tr>
<tr>
<td>5 degrees</td>
<td>720</td>
<td>731</td>
<td>723</td>
<td>631</td>
</tr>
<tr>
<td></td>
<td>(.07)</td>
<td>(.06)</td>
<td>(.07)</td>
<td>(.05)</td>
</tr>
<tr>
<td>10 degrees</td>
<td>706</td>
<td>733</td>
<td>724</td>
<td>667</td>
</tr>
<tr>
<td></td>
<td>(.09)</td>
<td>(.11)</td>
<td>(.09)</td>
<td>(.13)</td>
</tr>
</tbody>
</table>

Discussion.
the parafoveal stimulus, so that a 5 degree parafoveal preview was more useful than a 10 degree preview, \( F(1,7) = 36.4570, p < .001 \), for the interaction of Parafoveal Preview by Eccentricity. At 5 degrees, mean naming latency was 725 msec without a preview and 622 msec with a preview, indicating a preview benefit of 103 msec. At 10 degrees, naming latencies were 721 msec without a preview and 673 msec with a preview, giving a benefit of 48 msec. The main effect of Eccentricity was also significant by itself \( F(1,7) = 34.5474, p < .001 \), although this effect was clearly mediated by the interaction with Parafoveal Preview.

The effect of Foveal Prime was also significant \( F(2,14) = 3.8981, p < .05 \). Mean naming latencies for the related, unrelated, and blob conditions were 681, 694, and 680 msec, respectively. Even though subjects were never explicitly told to attend to the foveal primes, and in fact were told to move their eyes as quickly as possible to the parafoveal stimulus, the foveal primes were encoded to a level where they could exert an influence on subsequent processing. Target naming latencies were facilitated when the foveal prime was related to that target compared to when the foveal prime was an unrelated object. This result suggests an explanation for at least some of the effect of context found in studies in which the likelihood of an object's
occurring in a scene is manipulated: the previously
fixed object affects the identification of the
currently fixated object.

Another important aspect of these data is that the
unrelated condition shows inhibition in relation to the
blob condition, while the related and blob conditions are
virtually identical. According to Posner and Snyder
(1975) two-process account of priming, a finding of
inhibition for unrelated primes in relation to a neutral
baseline indicates the use of an attentional process,
rather than the use of an automatic process such as
spreading activation. In other words, the fact that
inhibition was apparently dominant may indicate that
subjects were using attentional strategies, such as an
active prediction strategy, and were incurring a cost
when their expectations were violated. While this is a
possibility that cannot be ruled out in this experiment,
there are several aspects of the data that are
inconsistent with this interpretation.

First, an attentional expectancy strategy would
predict not only a cost for trials on which the
prediction was incorrect, but also some facilitation for
those trials on which the prediction turned out to be
correct, such as on the related-prime trials in this
experiment. However, the related and blob conditions were
virtually identical, 681 versus 680 msec, making it seem
unlikely that a prediction strategy was being employed. Second, the mean saccade latency in this experiment was 250 msec, meaning that the stimulus onset asynchrony (SOA) between a foveal fixation of the prime and the target was about 285 to 300 msec (250 msec saccade latency plus approximately 35 to 50 msec saccade duration). In previous experiments which used the picture priming paradigm (e.g. Carr et al., 1982; McCauley et al., 1980), SOAs of over 500 msec were found to produce automatic facilitation. Therefore, 300 msec is probably too little time for attentional processes to become active (see also Neely, 1977). Finally, evidence to be presented in Experiment 2 is inconsistent with this interpretation.

There is an alternative to the hypothesis that the use of an attentional strategy caused the inhibition shown for the unrelated condition. The choice of an appropriate neutral baseline has been an issue in the word-priming literature (e.g. deGroot, 1983), and a similar issue can be raised here. In particular, the blob chosen in the current experiment as the neutral stimulus may, in hindsight, have been a poor choice. First, the blob, unlike the related and unrelated primes, did not have a name or label associated with it, at least not one that would be automatically activated, as would have been the case with the objects. Therefore, while
there was the possibility for Stroop-like name competition between the prime and target in the case of the object primes, no such possibility existed for the blob prime. Naming latencies with the blob prime may have been speeded due to the lack of this type of name competition. Further, aside from the problem of name competition in the case of the object primes, McCauley et al. (1980) have argued that the process of attaching a name to a priming picture during an object priming paradigm is a capacity demanding operation that can interfere with the naming of a subsequent picture. Therefore, covertly naming the foveal prime, holding the name in an STM buffer, and then preparing a different name as a response may all tend to slow the object prime conditions in relation to the blob prime condition.

Second, while the neutral stimulus was equated with the objects for the average number of pixels used to create the stimuli, it may have been that the actual visual complexity of the two types of primes was not equated, such that the object primes were more visually complex. The more complex object primes may have required more visual processing capacity for analysis. Further, meaningful visual stimuli may also capture more central processing capacity, since a meaningful stimulus will tend to make contact with long-term memory
representations. This would again leave less capacity for processing the parafoveal stimulus in the case of the related and unrelated primes, while nearly all capacity could be used for processing the parafoveal preview and the target stimulus in the case of the non-meaningful blob prime. All or any of these differences between the object primes and the neutral prime would tend to have the effect of underestimating the response time in the neutral condition. An underestimated neutral response time, in turn, would have the effect of underestimating facilitation and overestimating inhibition.

Table 3 presents the interaction of Foveal Prime with Parafoveal Preview [F(2, 14) = 5.6000, p < .05]. As can be seen in Table 3, this interaction results from the greater facilitative effect of the parafoveal preview on naming latency when there was a blob in the fovea compared with when there was a related object in the fovea. Removing the blob condition from the analysis caused the interaction to disappear, F(1, 7) = 4.0557, p > .05. It appears as though the subjects were better able to make use of the parafoveal information when there was a non-meaningful object in the fovea. This lends credence to the view that the blob condition was not really neutral, but instead engaged fewer processing resources and therefore underestimated the true baseline.
Table 3

Mean naming latencies for the three levels of Foveal Prime and two levels of Parafoveal Preview. Also shown is the relative amount of benefit due to the preview.

<table>
<thead>
<tr>
<th>Foveal Prime</th>
<th>Parafoveal Preview</th>
<th>rel</th>
<th>unrel</th>
<th>blob</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Preview</td>
<td></td>
<td>713</td>
<td>732</td>
<td>724</td>
</tr>
<tr>
<td>Preview</td>
<td></td>
<td>649</td>
<td>656</td>
<td>637</td>
</tr>
<tr>
<td>Benefit</td>
<td></td>
<td>64</td>
<td>76</td>
<td>86</td>
</tr>
</tbody>
</table>

Table 4 presents the interaction of Foveal Prime with Eccentricity \([F(2,14) = 8.0359, \ p < .005]\). It is clear from Table 4 that moving the target closer proved more facilitative when there was a blob in the fovea than when there was an object in the fovea. It is tempting to conclude again that it is easier to make use of the closer parafoveal information when there is a blob rather than an object in the fovea. However, this time the story is not quite as clear as it was in the case of the Foveal Prime by Parafoveal Preview interaction, since: (1) The difference in the "benefit" (10 degrees
minus 5 degrees) between the blob and unrelated condition is small and not significant by a simple effects analysis ($F < 1$), and (2) The Eccentricity factor does not distinguish between presence or absence of a parafoveal preview. If the explanation for the relatively greater facilitation for a closer target with a blob in the fovea is that the blob does not interfere as much with parafoveal processing (which is more important with a closer parafoveal preview) as do the object primes (an argument similar to the one made above for the Foveal Prime by Parafoveal Preview interaction), then the relatively greater facilitation found with the blob in the fovea and the closer target should only have occurred when there was, in fact, a parafoveal preview. While the three-way interaction between Foveal Prime, Parafoveal Preview, and Eccentricity did not approach significance ($F < 1$), the data clearly show a trend in the direction suggested, as can be seen in Table 5: Moving the target closer was more beneficial with a blob rather than an object in the fovea mainly when there was also a preview in the parafovea. While it would be unwise to make too much out of a nonsignificant interaction, the data in Table 5 are suggestive of the possibility that the blob condition was deterring less from the ability to process a parafoveal stimulus than were the object prime conditions, leading again to the conclusion that placing
a non-nameable, non-meaningful blob, rather than a nameable, meaningful object, in the fovea makes parafoveal information extraction easier.

Table 4

Mean naming latencies for the three levels of Foveal Prime and two levels of Eccentricity. Also shown is the relative amount of benefit due to the closer eccentricity.

<table>
<thead>
<tr>
<th>Eccentricity</th>
<th>rel</th>
<th>unrel</th>
<th>blob</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 Degrees</td>
<td>686</td>
<td>708</td>
<td>697</td>
</tr>
<tr>
<td>5 Degrees</td>
<td>675</td>
<td>680</td>
<td>664</td>
</tr>
<tr>
<td>Benefit</td>
<td>11</td>
<td>28</td>
<td>32</td>
</tr>
</tbody>
</table>

In sum, it is not possible from these data to determine the exact nature of the difference between having the objects versus the blob in the fovea. It seems likely that the blob was processed differently in at least two respects: First, the blob allowed more information to be picked up from the parafovea when that
information was available. This is interesting in its own right, since it suggests that the amount of parafoveal information picked up on a fixation depends upon the complexity of the object in the fovea. Second, the blob disrupted the naming of the target less, possibly because there was no chance for either name competition or confusion at the response stage. Given these differences, any analysis using the blob as a baseline from which to assess priming would seem unwise.

Table 5
Mean naming latencies (in msec) with an object versus a blob in the fovea, for the two levels of Parafoveal Preview and two levels of Eccentricity. Also shown is the relative amount of benefit due to the closer eccentricity.

<table>
<thead>
<tr>
<th>Eccentricity</th>
<th>No Preview</th>
<th>Preview</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>object</td>
<td>blob</td>
</tr>
<tr>
<td>10 Degrees</td>
<td>719</td>
<td>724</td>
</tr>
<tr>
<td>5 Degrees</td>
<td>725</td>
<td>723</td>
</tr>
<tr>
<td>Benefit</td>
<td>-6</td>
<td>1</td>
</tr>
</tbody>
</table>
Since the neutral foveal prime condition in this study seems suspect, and since Experiment 2 attempts to provide a more judiciously chosen neutral baseline, "priming" in Experiment 1 will henceforth be examined in terms of the differences in naming latency between the related and unrelated foveal prime conditions rather than in relation to the blob condition.

**Analysis Without Baseline.** A second ANOVA, identical to the first but excluding the blob condition, was conducted in order to examine the effects of the related and unrelated foveal primes. Only those results related to the foveal prime conditions will be reported, though all other effects mirrored those reported above in the overall analysis.

This analysis revealed that the object levels (related versus unrelated) of the factor Foveal Prime differed significantly from each other \( F(1,7) = 6.3205, p < .05 \). There was an overall priming effect of 13 msec for related vs unrelated primes. These two levels of Foveal Prime also interacted with Eccentricity \( F(1,7) = 7.9176, p < .05 \). At 10 degrees, there was a 22 msec priming effect, while at 5 degrees the effect was only 5 msec. The two levels of Foveal Prime also interacted marginally with Parafoveal Preview \( F(1,7) = 4.2234, .05 < p < .10 \), such that the priming effect was 19 msec with
no preview and 7 msec with a parafoveal preview. The effects of Parafoveal Preview and Eccentricity were additive with respect to the priming effect ([F < 1] for the three way interaction of Foveal Prime X Parafoveal Preview X Eccentricity). These interactions can be seen in Table 6 where the data are expressed in terms the priming effect -- the differences in latency between the related and unrelated foveal prime conditions.

As can be seen in Table 6, the related foveal prime was most useful when there was no preview and the target was furthest away, at 10 degrees. Less facilitation was found when there was a preview or when the target was closer, and finally no facilitation was found at all for the related over the unrelated prime when the preview appeared at 5 degrees in the parafovea.

Table 6

Amount of priming (unrelated minus related conditions) in msec by Parafoveal Preview and Eccentricity.

<table>
<thead>
<tr>
<th>Eccentricity</th>
<th>No Preview</th>
<th>Preview</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 degrees</td>
<td>11</td>
<td>-2</td>
</tr>
<tr>
<td>10 degrees</td>
<td>27</td>
<td>16</td>
</tr>
</tbody>
</table>
These results show that the related foveal prime was useful under certain circumstances. In particular, it appears that the related prime was most useful when the target was difficult or impossible to see in the parafovea, i.e. the 10 degree eccentricity and no preview conditions, and least useful when the target could be processed easily in the parafovea, i.e. when there was a preview at 5 degrees. In general, then, it appears that priming occurs only when the target object cannot be easily processed in the parafovea.

This general conclusion cannot be entirely complete, however: Another interesting aspect of the data presented in Table 6 is that there was a difference in the amount of priming found at 5 and 10 degrees, even when there was no parafoveal preview. This difference is somewhat surprising; eccentricity here refers only to the distance the eye had to travel in order to fixate the eventual target, and not to the distance of a parafoveal preview, since in this condition there was no parafoveal preview. In other words, these conditions were virtually identical aside from the distance that the eye had to travel to fixate the target. Why, then should there be a difference in priming of this magnitude between these two cells?

For the moment, this issue will be put off, and will be taken up again in Chapter 3.
In conclusion, several general statements about the data from this experiment can be made. First, it is clear that visual information about an object gathered from the parafovea aids subsequent identification of the object when that object is fixated. This replicates the work of Pollatsek et al. (1984) and extends it to a situation in which there is a meaningful object in the fovea. It appears that while more information can be extracted from the parafovea when there is a non-meaningful stimulus in the fovea, a great deal can also be extracted when there is a meaningful object in the fovea, even out to 10 degrees visual angle (see Table 7).

Table 7

Parafoveal Preview effect (no preview minus preview) as a function of Eccentricity and type of foveal stimulus.

<table>
<thead>
<tr>
<th>Eccentricity</th>
<th>Foveal Stimulus</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Object</td>
</tr>
<tr>
<td>5 Degrees</td>
<td>95</td>
</tr>
<tr>
<td>10 Degrees</td>
<td>44</td>
</tr>
</tbody>
</table>

Second, identification of a fixated object is
affected by the object fixated immediately before. Particularly, an object is identified faster if the object seen on the previous fixation was related rather than unrelated to it. This aspect of the data thus supports the intralevel priming model of context effects in scene perception.

Finally, recall that it was originally hypothesized that the foveal primes and parafoveal preview might show an over-additive relationship, such that there would be more facilitation from a parafoveal preview when there was a related compared to an unrelated object in the fovea. Such a result would have indicated that parafoveal information was more useful given a related foveal object. Instead, the marginal interaction between these two factors is in the opposite direction. It appears that, if anything, the parafoveal preview was less useful given a related foveal prime; or, to turn it around, the related foveal prime was less useful given a parafoveal preview. If this interaction is reliable, it indicates that priming of the sort shown here is useful only when the to-be identified object is difficult to see, for example when it is far away or when it is masked by other objects.

Experiment 2 seeks to replicate and extend these results.
Experiment 2

In Experiment 1 it was shown that the identification of an object can be facilitated if a related object rather than an unrelated object is viewed on the previous fixation. However, since a meaningless blob was used as the control prime, it was impossible to determine whether the difference between the related and unrelated primes was due to actual facilitation from the related object, inhibition from the unrelated object, or some combination of both. The distinction between facilitation and inhibition is theoretically important, since the automatic priming process posited here as an account of the context effects found in scene processing specifically predicts that facilitation without inhibition should be found. On the other hand, if the priming effect observed in Experiment 1 was due to an expectancy strategy, where subjects allocated attention to a particular response given a particular prime, then inhibition would be predicted when the target was not the expected object.

In order to determine whether the priming effect demonstrated in Experiment 1 was facilitation rather than inhibition dominant, a different and more diagnostic neutral prime was chosen for the current experiment. Specifically, four objects which were not
predictive of any of the thirty targets were chosen to serve as neutral primes. These four neutral primes appeared randomly whenever a neutral prime was called for. Therefore, there was no way for subjects to use an expectancy strategy when these objects appeared. If the priming effect found in Experiment 1 was due to facilitation without inhibition, then the related prime condition in the current experiment should be faster than both the unrelated and neutral prime conditions, while the latter two should not differ from each other. If, on the other hand, the priming effect is due to non-automatic processes involving allocation of attention, then the unrelated prime condition should show inhibition in relation to the neutral prime condition.

A second purpose of the current experiment was to determine whether a priming effect could be made to occur at 5 degrees even when the subject was allowed a parafoveal preview of the target, if the parafoveal preview were made more difficult to see. (Recall that in Experiment 1, there was a tendency for the priming effect to be smaller or even disappear if the target could be seen clearly in the parafovea, that is, if there was a close parafoveal preview.) One of the differences between normal scenes and the stimuli used in Experiment 1 is that in scenes, parafoveal objects are surrounded by other objects and background, making them more difficult
to see. In order to simulate this in the paradigm used here without adding the confound of having two nameable objects in the parafovea, the same blob as was used in Experiment 1 was placed between the foveal prime and the parafoveal preview on half of the trials so that the preview would be more difficult to see. It was expected that this would increase the amount of priming shown at 5 degrees.

Finally, some of the results found in Experiment 1 were unexpected. For example, more priming was found at 10 degrees than at 5 degrees even when there was no parafoveal preview of the target. It is not entirely clear why this should be so. Further, there was a tendency for there to be less distance and preview benefit for related compared with unrelated foveal primes. Experiment 2 served to determine whether these results were replicable.

Method

Subjects. Eight members of the University of Massachusetts subject pool participated in the experiment. Six of the eight subjects had also participated in Experiment 1.

Materials. The stimuli used were the same 60 line drawings as used in Experiment 1. In addition, the blob used as a foveal prime in Experiment 1 was used here as a
parafoveal "mask". Also, four new line drawings taken from Snodgrass & Vanderwart (1980) replaced the blob as the neutral prime. These objects were a bed, a cannon, a snowman, and a stoplight. It should be noted that each of these neutral primes was in fact related to at least one of the targets in some way. This could not be helped, since with thirty targets, it is virtually impossible to find 4 objects which are totally unrelated to any of the targets. The important point to keep in mind, however, is that, given one of the 30 non-neutral primes, there was a .50 probability that a particular related object would be the target and a .50 probability that a particular unrelated object would be the target, while given one of the neutral primes, the probability was only .033 that the quasi-related object would be the target, and the probability was also .033 that the target would be any particular other target object. Therefore, with the neutral primes, the probability was .966 that the target would not be the quasi-related object, and so the target was virtually unpredictable given a neutral prime.

As in Experiment 1, subjects were asked to name each of the objects before the experiment, and were corrected until they had the appropriate name for each object.

Apparatus. The equipment used was the same as used in Experiment 1. Aside from all of the functions already described above, the computer also randomized which of
the four neutral objects would occur on a particular neutral trial.

Procedure. The procedure was the same as in Experiment 1, with the following exceptions. First, as already described, the neutral foveal prime condition consisted of four objects rather than the meaningless blob. On a neutral trial, the computer randomly selected which one of the four neutral objects would be displayed.

Second, a new factor, Parafoveal Mask, was introduced, which was fully crossed within subject with all other factors. The parafoveal mask consisted of the blob used in Experiment 1. On half of the trials, this lateral mask appeared spatially between the foveal prime and the parafoveal stimulus (target or box). On the other half of the trials, the mask did not appear. When the mask did appear, its nearest outer edge was 1/2 degree from the nearest outer edge of the parafoveal stimulus. If the mask was on at the start of a given trial, it remained on after the eye movement as well.

Like the factor Eye Movement Direction, Parafoveal Mask was blocked. Therefore, all subjects participated in four blocks, all possible combinations of Parafoveal Mask (mask or no mask) and Eye Movement Direction (left-to-right or right-to-left). The order of blocks was counterbalanced across subjects according to a Latin Square. The experiment was completed in four sessions,
one session for each block, generally run on consecutive
days, with each session lasting 45 to 60 minutes.

Design

The experiment consisted of 1440 trials per
subject. The 1440 trials consisted of the factorial
combination of 30 (parafoveal targets) X 3 (related
versus unrelated versus neutral foveal prime) X 2
(parafoveal mask versus no parafoveal mask) X 2
(parafoveal preview versus no parafoveal preview) X 2
(5 degrees versus 10 degrees visual eccentricity) X 2
(left versus right eye movement direction). Eye Movement
Direction and Parafoveal Mask were blocked and
counterbalanced across subjects according to a Latin
Square.

Results and Discussion

The corrected mean naming latencies, collapsed over
items, subjects, direction of eye movement, and
parafoveal mask, are presented in Table 8. Naming errors
were again very infrequent (occurring on less than .01 of
the trials) and were randomly distributed across
conditions. As in Experiment 1, the analyses reported
here were conducted on the corrected mean response
times, with noise trials excluded. The mean percentage of
noise trials for each condition are also shown in
Table 8. The pattern of results for the corrected mean naming latencies did not differ from the pattern before correction.

Table 8
Mean time to name the target object in msec (and mean percentage of noise trials) by Eccentricity, Parafoveal Preview, and Foveal Prime.

<table>
<thead>
<tr>
<th>Eccentricity</th>
<th>No Preview</th>
<th></th>
<th></th>
<th>Preview</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>rel</td>
<td>unrel</td>
<td>blob</td>
<td>rel</td>
<td>unrel</td>
<td>blob</td>
</tr>
<tr>
<td>5 degrees</td>
<td></td>
<td></td>
<td></td>
<td>708</td>
<td>712</td>
<td>716</td>
</tr>
<tr>
<td></td>
<td>(.05)</td>
<td>(.07)</td>
<td>(.06)</td>
<td>607</td>
<td>606</td>
<td>606</td>
</tr>
<tr>
<td>10 degrees</td>
<td></td>
<td></td>
<td></td>
<td>708</td>
<td>725</td>
<td>726</td>
</tr>
<tr>
<td></td>
<td>(.07)</td>
<td>(.09)</td>
<td>(.08)</td>
<td>658</td>
<td>670</td>
<td>675</td>
</tr>
</tbody>
</table>

A 3 (Foveal Prime) X 2 (Parafoveal Mask) X 2 (Parafoveal Preview) X 2 (Eccentricity) X 2 (Eye Movement Direction) within-subject ANOVA was conducted on the mean naming latencies, treating subjects as a random effect.

The first thing to note about the data from this experiment is that the manipulation of Parafoveal Mask
was totally without effect. Placing the non-meaningful blob between the foveal prime and parafoveal stimulus produced neither a main effect \((F < 1)\) nor interacted with any other factor in any interpretable way. (Parafoveal Mask did participate in two higher order interactions, Foveal Prime by Parafoveal Mask by Parafoveal Preview by Eye Movement Direction \([F(2,14) = 6.2821, p < .05]\) and the five-way interaction involving all factors \([F(2,14) = 4.8559, p < .05]\), but neither of these interactions allowed any obvious interpretation. Given the number of factors in the current design, and an overall alpha level of .05, one or two spurious effects would be expected. Therefore, these interactions will be ignored.) Of particular importance is the fact that the presence of the parafoveal mask did not increase the amount of priming found at 5 degrees when there was a parafoveal preview (3 msec priming without the mask, -6 msec priming with the mask, neither of which differed from 0 by \(t\) test), as would be predicted if the lack of a priming effect at 5 degrees with a preview were due to the ease of seeing the preview. However, since there was no overall effect of the parafoveal mask, it appears that subjects were able simply to ignore it, and therefore this condition does not allow a test of the hypothesis that priming would be found at 5 degrees if the parafoveal preview were made more difficult to see.
As in Experiment 1, Direction of Eye Movement again produced no main effect \( (F < 1) \), though it did participate in the two higher order interactions described above. However, as those interactions had no apparent meaning, this factor will not be discussed further.

As would be expected, given the results of Pollatsek et al. (1984) and Experiment 1 above, there was a main effect of Parafoveal Preview \( [F(1,7) = 207.0796, p < .00005] \). The mean naming latencies with and without a parafoveal preview were 637 and 716 respectively. In addition, there was again a significant Parafoveal Preview by Eccentricity interaction, showing more of a parafoveal benefit at a 5 degree eccentricity than at 10 degrees \( [F(1,7) = 71.1612, p < .0005] \). At 5 degrees, mean naming latency was 712 msec without a preview and 606 msec with a preview, giving a preview benefit of 106 msec (compared with 103 msec benefit in Experiment 1). At 10 degrees, naming latencies were 720 msec without a preview and 668 msec with a preview, for a preview benefit of 52 msec (compared with 48 msec in Experiment 1). The significant main effect of Eccentricity \( [F(1,7) = 62.0585, p < .0005] \) was clearly primarily due to the trials on which there was a preview. These results unambiguously show that subjects are able to use parafoveal information quite far in the periphery.
Of primary interest in the current experiment is the effect of the type of foveal prime seen on a trial. Consistent with the view that context effects in scene processing can be accounted for through the operation of passive spreading activation (intra-level priming), the main effect of Foveal Prime was significant \(F(2,14) = 9.3399, p < .005\). Mean naming latencies were 670, 678, and 681 msec for the related, unrelated, and neutral prime conditions, respectively. Planned comparisons showed that the difference between the related and unrelated conditions \(F(1,7) = 8.2363, p < .05\) and the difference between the related and neutral conditions \(F(1,7) = 11.3216, p < .05\) were both significant, while the difference between the unrelated and neutral conditions was not \(F(1,7) = 3.5648, p > .05\). Using the neutral condition as a baseline, it appears that there is an overall facilitation effect of 11 msec for a related prime, and no cost for an unrelated prime. Therefore, within the Posner and Snyder (1975) framework, these results indicate automatic facilitative processing.

Finally, the pattern of results observed in Experiment 1 (when the non-meaningful blob prime was removed from the analysis) between Foveal Prime, Eccentricity, and Parafoveal Preview, was again found in Experiment 2. Table 9 presents the the amount of priming
found (the difference between the neutral condition and the related and unrelated conditions) as a function of Eccentricity and Parafoveal Preview. First, the interaction of Foveal Prime with Eccentricity was significant \( F(2,14) = 6.3801, p < .05 \), showing more of a priming effect at 10 degrees (18 msec) than at 5 degrees (3 msec). Second, there was again a moderate though non-significant tendency for there to be a larger priming effect when there was no parafoveal preview (13 msec) compared with when there was a preview (7 msec) \( F(2,14) = 2.5900, p = .11 \). This result is inconsistent with the hypothesis that more parafoveal information can be extracted with a related object in the fovea. Third, the three-way interaction between these factors was not significant \( F < 1 \), indicating that the eccentricity benefit on priming was as large when there was no preview as when there was a preview. As was indicated in discussing Experiment 1, this last result is somewhat counter-intuitive, since when there was no preview, the only difference between the 5 degree and 10 degree eccentricity conditions was the distance the eye had to travel, and it is not clear why travelling further should increase the amount of priming observed. Several hypotheses will be considered as explanations for this effect in Chapter 3.
In summary, Experiment 2 has replicated and extended the results of Experiment 1. A consistent priming effect was again found, such that the identification of an object fixated after an eye movement was facilitated given that a related object had been previously fixated. In addition, no difference in target identification time was found depending on the predictability of the target given an unrelated object (related versus neutral primes). Therefore, these data are consistent with an automatic and passive spreading activation account of the priming effect. An attempt was made to determine whether the priming effect would be increased if the parafoveal preview of the target...
were made less salient. However, the manipulation introduced to test this was unsuccessful in that it produced no effect at all. Therefore, whether a priming effect will occur at close eccentricities if the target object is less easily seen in the parafovea remains an open issue.

The general effect of allowing subjects a parafoveal preview of the target object was replicated. Identification time was facilitated given a parafoveal preview, and this effect was greater if the preview was closer. The benefit of having the target in the parafovea or nearby was also again found to be contingent on the type of foveal prime presented: There was less distance and preview benefit given a related prime.

In the next chapter, several additional analyses will be presented. These analyses attempt to rule out some uninteresting explanations of these effects and further try to narrow down the exact causes of the patterns of data observed. In addition, a model will be presented which accounts for the major trends in the data and which suggests possible future directions.
In this chapter several additional analyses are discussed which attempt to narrow down possible interpretations of the data presented in Chapter II.

Priming, Preview, and Eccentricity

Recall that in Experiments 1 and 2, more priming was found at 10 degrees than at 5 degrees, even when there was no parafoveal stimulus. In this section, several explanations for this anomalous finding are examined.

One reasonable explanation for the greater amount of priming at 10 degrees than at 5 degrees even when there was no preview is suggested by differences in saccade latency and saccade duration. An examination of the saccade latency data (where saccade latency is defined as the amount of time it took to begin an eye movement toward the parafoveal target once the trial display appeared) showed that there was a significant effect of eccentricity, such that it took an average of 240 msec versus 260 msec to initiate a saccade to a 5 degree versus 10 degree target in Experiment 1 \(E(1,7) = 113.5447, p < .0001\), and 230 msec versus 244 msec in Experiment 2 \(E(1,7) = 43.7467, p < .001\). Differences in saccade latencies of this type are not unusual.
(e.g. Pollatsek et al., 1984), and are partially the result of the longer time needed to adequately locate the parafoveal stimulus, as well as the longer time needed to program the saccade at 10 degrees. In addition, saccade durations (the amount of time the eye spends travelling) are typically longer at greater eccentricities. If the assumption is made that activation can build to a higher level at the target node given a longer stimulus onset asynchrony (SOA), then an explanation for the priming difference at 10 degrees versus 5 degrees might be that at 10 degrees the longer saccade latency and saccade duration allows more activation to build up, and so produces more priming.

If the longer SOA at a 10 degree eccentricity is the cause for the greater amount of priming found at 10 degrees, then more priming should generally be found when the saccade latency is increased. In order to test this assertion, the data from both experiments were divided in a mean split according to saccade latency. The mean saccade latency for each subject in each block at each eccentricity was computed, and the mean naming latency was found for those trials on which saccade latency was below the mean saccade latency and those trials on which saccade latency was above the mean. These data were then subjected to a 5-way ANOVA for the Experiment 1 data and a 6-way ANOVA for the Experiment 2 data, with Saccade
Latency added to the original factors. There was, in fact, no tendency for there to be more priming at slower saccade latencies, either overall (Foveal Prime X Saccade Latency, Experiment 1, $[F < 1]$; Experiment 2, $[F(2,14) = 1.0053]$), or by eccentricity (Foveal Prime X Saccade Latency X Eccentricity, Experiment 1, $[F < 1]$; Experiment 2, $[F(2,14) = 3.0891, .05 < p < .10]$) In fact, this last marginal interaction actually went in the wrong direction, showing more priming at 5 degrees with a shorter saccade latency. Thus, the data do not support the idea that the longer saccade latencies at 10 degrees are the cause of the greater priming found at 10 degrees in the no preview condition.

A second possible explanation for this effect is that at 5 degrees, the box-with-cross used as the parafoveal target in the no preview condition was encoded to a degree sufficient to cause disruption to the process which integrates information across saccades. As has been shown in the current experiments as well as by Pollatsek et al. (1984; in progress), a great deal of parafoveal information is picked up at 5 degrees. Assuming that an inconsistency between information acquired in the parafovea (a box) and information acquired when that area is fixated (an object) causes disruption of the integration process (and possibly causes the visual system to adopt an unusual processing mode), the finding
of less priming in the no preview condition at 5 degrees could be the result of an artifact of the present experimental setting. Unfortunately, there is no obvious way to directly test this possibility given the data at hand.

A third possible explanation for the equivalent amount of priming found at 5 degrees and 10 degrees given no parafoveal preview takes into account the fact that the accuracy of a saccade depends, in part, on the distance the eye has to travel. If subjects tend to land more accurately on the target object given a shorter saccade, then the greater amount of priming found at 10 degrees may be due to the fact that the quality of the visual information picked up after a 10 degree saccade is less than the quality picked up after a 5 degree saccade. Previous work has shown that priming effects increase when the target is visually degraded (Meyer et al, 1975; Sperber et al, 1979). Therefore, if subjects do tend to land less accurately on the targets at 10 degrees, a likely explanation for the anomalous effect would be possible.

In order to determine whether subjects were in fact more accurate when moving their eyes 5 degrees rather than 10 degrees, two indices of accuracy were computed. The first index compared fixation position upon first landing with fixation position when the target was
named. Presumably the eye was closer to the preferred viewing location when the target was named, and so the difference between the fixation position when the target was named (Fix 2) and the fixation position when the eye first landed (Fix 1) should give an indication of how close to the preferred viewing location the first landing was. The second index was the standard deviation of the first fixation position. The more variable the fixation position, the less accurate it would be on the average. Both of these measures were computed in terms of pixels, where 18 pixels equalled 1 visual degree.

In both Experiment 1 and Experiment 2, the above indices provided evidence that the accuracy of the saccades were inversely related to the eccentricity of the target. In Experiment 1, the mean difference between Fix 1 and Fix 2 was 12 pixels at 5 degrees and 31 pixels at 10 degrees [$t(7) = 4.1319, \ p < .005$], and the mean standard deviation of Fix 1 was 19 pixels at 5 degrees and 52 pixels at 10 degrees [$t(7) = 6.0314, \ p < .001$]. Similarly, in Experiment 2 the mean difference was 9 at 5 degrees and 19 at 10 degrees [$t(7) = 5.5961, \ p < .001$], and the mean standard deviation was 16 at 5 degrees and 35 at 10 degrees [$t(7) = 4.2817, \ p < .005$]. Thus, support is lent to the notion that a difference in the accuracy of the saccade accounts for the differential priming found at 5 degrees and 10 degrees when there was
no parafoveal preview. In addition, the finding presented above that there may be more priming given a shorter saccade latency is also consistent with this interpretation if it is assumed that fast saccades are also less accurate.

A more direct test of the "bad landing gives more priming" idea would be to examine only those trials on which the subject landed accurately at 10 degrees, and then to determine whether there was less priming on these trials. In order to attempt such a test, the mean Fix 2 (preferred viewing location) for each subject in each block at each eccentricity was computed. Naming latencies were then computed for only those trials on which Fix 1 (where the subject first landed) was within 1 degree to either side of this position. Since the target objects were 2 degrees horizontally, this would ensure that Fix 1 was on the object. As expected, many more trials met this criterion at 5 degrees than at 10 degrees. However, since there were so few trials which met the criterion at 10 degrees, the data were extremely variable and no direct test was possible. Therefore, while the data support the poorer accuracy of a 10 degree saccade, it must be left to future research to determine whether this does in fact increase the effect of context on object identification.
Items Analysis

In order for the intralevel priming explanation of context effects in scene viewing to be a viable alternative to schema theory, it is important that the priming effect shown in the last chapter not be the result of a few of the object pairs used as stimuli. Therefore, items analyses were conducted. These analyses showed exactly the same patterns of significance as were found treating subjects as the random effect, both in Experiment 1 and Experiment 2. Therefore, there is no evidence that a few items were producing the effects found.

Name Frequency Analysis

The naming paradigm employed in this thesis involves at least two separate stages of processing, an object identification stage and a name retrieval/production stage. Therefore, it is possible that the demonstrated priming effect had its locus at either of these two stages. Since name retrieval is not a logically necessary stage in normal object identification, the generality of the priming effect to scenes would be reduced if it were the case that the priming effect were occurring predominantly at the name retrieval/production stage. In order to test whether the priming effect was occurring at the latter stage, additive factors logic (Sternberg,
1969) was employed. If the priming effect was occurring at the name retrieval/production stage, then priming should interact with another factor known to affect this stage. Such a factor is the frequency in the language of the word or name produced. If, on the other hand, the priming effect is occurring at the object recognition stage, then the effects of name frequency and prime should combine additively.

In order to test this, the name frequencies of the 30 target objects were found in Kucera and Francis (1967). The targets were then rank ordered and split into 3 groups of 10 according to name frequency. An ANOVA for each experiment was then conducted on the mean naming latencies, averaged over subjects, treating Name Frequency as a between items factor. The results of these analyses were clear: While there was a main effect of Name Frequency in both Experiment 1 \( F(2, 27) = 5.1672, p < .05 \) and Experiment 2 \( F(2, 27) = 5.1048, p < .05 \), such that objects with higher frequency names were named faster, there was no hint in either experiment of a Foveal Prime by Name Frequency interaction (both Fs < 1). Therefore, these results are consistent with the conclusion reached by previous researchers that the object priming effect is not due to object naming. It appears, instead, that the priming effect is located at the object recognition level of processing (Huttenlocher
Visual Similarity Analysis

Aside from a passive spreading activation account of the priming effect demonstrated here, it is possible that the facilitation found for related primes was due to the greater visual similarity of the related primes to the target objects. Subjectively, the related primes did seem to be more visually similar to the targets than the unrelated primes. A visually similar prime could facilitate the low level feature processing of the target through simple feature overlap (Sperber et al. 1979). A priming effect due to simple feature overlap between related objects would suggest an additional explanation of the context effects found in scenes (since related objects typically look more like each other, even in scenes). However, such an effect would seem less robust than a priming effect at the semantic level, since it would be affected by such visual stimulus factors as object orientation.

In order to determine whether visual similarity was playing a role in the priming effect shown here, 4 subjects were asked to rate the related pairs on a 5 point scale of visual similarity. The ratings were extremely reliable for the 10 least and 10 most visually
similar pairs, and these were selected as the most extreme test of the visual similarity hypothesis. The mean similarity rating for the low similarity group was 1.1 (range 1.0 to 1.25), and the mean for the high similarity group was 3.65 (range 3.0 to 4.5). The overall priming effect for these two groups of 10 items was then examined. As can be seen in Table 10, there is no indication of a reduced priming effect for the 10 targets which were less visually similar to their primes. This result is consistent with Huttenlocher and Kubicek (1983), who explicitly controlled the visual similarity of their related and unrelated primes to their

<table>
<thead>
<tr>
<th>Priming</th>
<th>All Items</th>
<th>High Sim</th>
<th>Low Sim</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>13</td>
<td>12</td>
<td>15</td>
</tr>
<tr>
<td>2</td>
<td>11</td>
<td>10</td>
<td>17</td>
</tr>
</tbody>
</table>

Table 10

The priming effect (in msec) in Experiments 1 and 2, for all targets (n=30) and for target which had high (n=10) and low (n=10) visually similar primes.
targets and still found a sizeable priming effect.

There is, therefore, no evidence that the priming effect found here can be explained at the level of visual similarity.

Naming Latency Frequency Distribution Analyses

The effect of having a parafoveal preview of the target was shown to be quite large and robust. The cause of this effect is thought to be an integration of the information picked up in the parafovea with the information picked up once the eye fixates the target (Pollatsek et al., 1984). In other words, since some information has been picked up in the parafovea, less processing needs to be done in order to identify the object once it has been fixated.

An alternative account of the preview effect suggests that subjects are sometimes identifying and beginning to name the parafoveal target before they move their eyes. According to this explanation, the preview effect is due to a full identification of the object in the parafovea on some proportion of the trials rather than to the integration of partial information across saccades. While Pollatsek et al. provided some evidence against the full parafoveal identification explanation, it would seem beneficial to show this for the current experiments. To this end, frequency distributions of the
naming latencies were prepared. If the parafoveal preview benefit is due to subjects identifying and beginning to name the target before they move their eyes on a significant proportion of the preview trials, then the naming latency distributions should tend to be bimodal when there is a preview. The two peaks of the bimodal distribution would reflect the trials on which subjects did and did not identify the parafoveal stimulus. On the other hand, if the parafoveal benefit is primarily due to the integration of information picked up before and after the eye movement, the naming latency distributions for the preview and no preview trials should be similar, with the mean of the former merely shifted to the left (faster responses).

In order to test these opposing predictions, frequency distributions of the naming latencies were created in the following manner: The mean naming latency for each subject in each experiment in each condition was found. Thus, these means collapsed over only the particular target presented. Next, the number of trials which fell into a distribution cell 25 msec wide was computed. These cells were centered at the mean naming latencies, i.e. the first cell greater than the mean would be from Mean to Mean + 25 msec, the next would be from Mean + 25 to Mean + 50 msec, etc., and the first cell less than the mean would be Mean - 25 msec,
etc. Next, these distributions were collapsed across eye movement direction, and the blocks which did not include the blob condition in Experiment 2 were collapsed with the Experiment 1 distributions. The collapsing was always centered at the means so that extraneous variables such as eye movement direction and practice effects across blocks would not affect the distributions.

The resultant distributions are shown depicted in Figures 2 through 5. (Only the related and unrelated foveal prime conditions are shown, since the blob prime trials of Experiment 1 could not be combined with the neutral prime trials of Experiment 2.) In essence, these figures are smoothed histograms, with the width of each histogram cell shrunk to a point. Each point along the X-axis therefore represents one distribution cell. For example, the point labeled "1" represents the cell from the Mean to the Mean + 25 msec, and the cell labeled "-1" represents the cell from the Mean to the Mean - 25 msec.

As can be seen in Figure 2 and Figure 3, the shapes of the distributions for the preview and no preview conditions are extremely similar. Figure 2 presents the preview and no preview distributions, centered around their means, for the related prime condition at a 5 degree eccentricity. Figure 3 presents the same distributions for the unrelated prime condition. While
the preview distributions are flatter and a bit wider than the no preview distributions, they are strikingly similar, and there is no evidence of bimodality given a preview.

Figure 4 and Figure 5 present distributions analogous to Figures 2 and 3, except with a 10 degree eccentricity. These distributions are more variable than their 5 degree counterparts, but the same conclusion emerges. There does not appear to be any evidence in these distributions favoring the hypothesis that the preview effect is due to the identification of the target in the parafovea. Instead, it appears that the parafoveal preview gives those targets a "head start", so that they are identified faster once they are fixated.

It should be pointed out that the slight flattening and bulging of the preview distributions (especially the bulges at about -200 msec from the mean) may indicate that on a small proportion of the trials subjects are in fact recognizing the target in the parafovea. The above discussion is not meant to suggest that this is not the case, especially when the preview is at 5 degrees. However, the important point is that the similarity of the distributions is substantial. Therefore, this analysis makes it appear extremely unlikely that such trials are the predominant cause of the preview effect.
Figure 2

Frequency Distributions for 5 degree related trials, preview and no preview, centered at their means.

5 Degrees, Related

![Graph showing frequency distributions for 5 degree related trials with and without preview, centered at their means.](image)
Figure 3

Frequency Distributions for 5 degree unrelated trials, preview and no preview, centered at their means.

5 Degrees, Unrelated

![Frequency Distributions Graph](image-url)
Figure 4

Frequency Distributions for 10 degree related trials, preview and no preview, centered at their means.

10 Degrees, Related

[Diagram showing frequency distributions for 10 degree related trials, with lines indicating 'no preview' and 'preview'.]
Figure 5

Frequency Distributions for 10 degree unrelated trials, preview and no preview, centered at their means.

10 Degrees, Unrelated
CHAPTER IV
DISCUSSION

In this chapter a model is presented which accounts for the experimental data. General conclusions are then discussed.

A Model

In this section a model will be proposed to account for the major trends in the data. It should be noted that this model is consistent with but not forced by these data. The model is meant to be a heuristic device and is speculative.

The model assumes a "pictogen" framework (Seymour, 1973; Warren & Morton, 1982) which is based upon but not identical to Morton's (1969) logogen model of word recognition. In this framework, each object will be thought of as having a categorizing element or pictogen which receives input from sensory and contextual sources, where contextual sources are defined in terms of objects only. Each pictogen has a resting level of activation. The level of activation can be increased above the resting level by input from the sensory and contextual inputs until it reaches a threshold, at which time the pictogen "fires" and the object is recognized.
The pictogen for an object will be thought of as having connections to the pictogens of other objects which are semantically related to it. For example, the pictogen representing a cat would have connections to the pictogens representing a dog, a lion, etc. Further, if the activation level of a pictogen reaches a high enough level, then this activation will spread to the other related pictogens, raising their activation levels also. Thus, encountering a cat will tend to raise the activation level of the pictogens of all related objects. However, there will be no effect on the pictogens for unrelated objects (see Posner & Snyder, 1975).

The outline above provides enough machinery for explaining the priming effect. It suggests that when a related prime is seen in the fovea, the level of activation for the corresponding pictogen increases, and this activation then spreads to the pictogens of related objects. When a related object is fixated next, it can be identified faster since its level of activation is already closer to the identification threshold. Less activation from the perceptual source is required in order to reach the threshold, so less time is spent processing the object perceptually.

In order to account for the parafoveal preview effect and its interaction with priming, an additional
assumption is necessary. This assumption is that the rate at which activation based on perceptual information accrues at the pictogen depends upon the quality of the visual stimulus. If the quality is good, then the activation level will increase quickly; if the quality is poor, then activation will increase more slowly. Such stimulus factors as eccentricity and preview will thus affect the rate: If the object is foveally fixated, then activation will rise rapidly. If the object is seen in the parafovea, but it is close by, then activation will rise more slowly. If the parafoveal preview is further away, then activation will rise more slowly still. Finally, if there is no preview, then there will be no perceptual input to the pictogen and the activation level will not change.

A general assumption of the model will be that the combined amount of activation which context and parafoveal inputs can produce is a parameter that changes with setting, task demands, etc. Under some conditions, subjects may allow these sources of activation to continue to produce activation until the identification threshold is reached. For example, if the task were parafoveal identification, then this might be the case. In applying the model to the experiments reported here, this assumption will be strengthened to state that the perceptual system puts a limit on the amount of
combined activation that it will allow the prime plus the parafoveal stimulus to produce. This strong assumption will be true whenever the object is going to be fixated; it prevents the prime and parafoveal glimpse from having too much weight and prevents misidentification. As has already been discussed, the data suggest that subjects were not identifying the parafoveal stimulus until it was fixated.

The experimental paradigm used in this thesis can be broken down into two separate segments of time. During the first, the preview period, the subject sees the foveal prime and the parafoveal stimulus. The foveal prime is encoded, and activation spreads to related pictogens, raising their activation levels. At the same time, if there is a parafoveal preview, activation from the perceptual source also begins to accrue at the pictogen of the target object, with the rate of accrual dependent on the eccentricity of the parafoveal preview. Activation continues to increase until either the eye moves or the limit is reached. Now the eye moves to the target object and the object is fixated. Perceptual activation increases in rate because the visual information is better, but there is still a difference in activation rate depending on the distance of the saccade. At 10 degrees, the eye lands less accurately, and activation accrues slightly less rapidly
than after a 5 degree saccade. In either case, identification takes place as soon as the activation level reaches the identification threshold.

This simple model accounts for all of the trends in the data. A numerical example will help make this clear. It should be kept in mind that the numbers chosen for the example are totally arbitrary. However, the pattern of numbers shows how the model works.

Assume that the resting level for any pictogen is 0 units, the recognition threshold is 100 units, and that a related prime adds 5 units of activation. Further, assume that if there is no parafoveal preview, 0 units are added. If there is a 10 degree preview, 2 units are added, and if there is a 5 degree preview, 10 units are added. Finally, assume that the limit that a prime plus a parafoveal preview can add is 10 units of activation.

When the display first appears on the screen, the prime is encoded, and if it is related, activation spreads to the target, raising its activation level by 5 units. At the same time, the parafoveal preview raises activation of the target as well. If the limit is reached, activation stops increasing. Table 11 displays the activation level for the target depending on the type of prime and the distance of the preview.
Table 11

The activation level of a pictogen depending on type of prime and parafoveal preview.

<table>
<thead>
<tr>
<th>Eccentricity</th>
<th>related</th>
<th>unrelated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>no prev</td>
<td>prev</td>
</tr>
<tr>
<td>5 degrees</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>10 degrees</td>
<td>5</td>
<td>7</td>
</tr>
</tbody>
</table>

Now the eye moves to fixate the target. If the eye moves 5 degrees, then assume that activation will increase at a rate of 15 units per unit time. If the eye moves 10 degrees, assume that the rate of activation increase is 10 units per unit time. The number of units of time that will be needed to reach the threshold will be \( \frac{(100 - \text{prior activation})}{\text{rate}} \). The number of units of time to reach identification threshold by condition is given in Table 12.

Table 12 represents the amount of time needed to identify the target in the various conditions. The numbers in this table are in arbitrary time units, and model only the identification time; later stages such as response selection and motor programming are thus not
modeled. The main effects of prime, preview, and eccentricity are evident, as well as the eccentricity by preview interaction.

Table 12

Length of time to reach threshold by condition.

<table>
<thead>
<tr>
<th>Eccentricity</th>
<th>Related</th>
<th>Unrelated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>no prev</td>
<td>prev</td>
</tr>
<tr>
<td>5 degrees</td>
<td>6.33</td>
<td>6.00</td>
</tr>
<tr>
<td>10 degrees</td>
<td>9.50</td>
<td>9.30</td>
</tr>
</tbody>
</table>

By subtracting the related times from the unrelated times at each level of eccentricity and preview, Table 13 is produced. As can be seen there, the eccentricity by prime, preview by prime, and even the anomalous greater priming at 10 degrees versus 5 degrees without a preview effect is evident. Finally, as was found in both experiments, no priming is produced at 5 degrees when there is a preview. A comparison of Table 13 with Table 9 from Experiment 2 shows remarkable similarity.
Table 13

The priming effect as predicted by the model.

<table>
<thead>
<tr>
<th>Preview</th>
<th>Eccentricity</th>
<th>none</th>
<th>target</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 degrees</td>
<td>0.33</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>10 degrees</td>
<td>0.50</td>
<td>0.50</td>
<td></td>
</tr>
</tbody>
</table>

The model thus does a good job of accounting for the data produced by the experiments reported in Chapter 2. It uses as a primary mechanism automatic spreading activation among nodes or pictogens in an organized semantic representation of objects. There seems to be no principled reason why the model cannot be extended to account for the effects of context found in scene perception also. Tests of the adequacy of the model in accounting for scene processing must await future research.
Conclusion

This thesis examined the effects of context on object identification. Numerous previous studies (discussed in Chapter 1) have shown that objects are more easily identified when they appear in an appropriate scene context than when they appear in an inappropriate context. Almost without exclusion, researchers working in this area have discussed these results in terms of a schema model. While these schema models have not been clearly spelled out, they all assume that a scene level representation is computed very rapidly (on the order of 100 msec), and that this representation then feeds information top-down to the object level, facilitating the recognition of objects which are predicted by the scene.

An alternative explanation for the effects of context on object identification was proposed in Chapter 1. This explanation, called the "intralevel priming" hypothesis, posits that some, if not all of the effects of context in scene processing, can be explained by passive spreading activation between nodes at the object level of representation. Consistent with this hypothesis, Experiment 1 demonstrated that objects were identified faster after a saccade if a related object rather than an unrelated object had been fixated
previously. Experiment 2 replicated this result and further showed that the benefit on identification time was due to facilitation from the related object rather than to inhibition from the unrelated object. Thus these data supported the hypothesis that passive spreading activation was producing the difference in identification time.

Aside from manipulating the relation of the object fixated prior to the fixation on the target object, these experiments also manipulated perceptual factors important in normal scene viewing. These factors included the distance of the target object prior to the saccade and whether or not the target object could be seen in the parafovea. Based on the results of these manipulations, in combination with the prime manipulation, a model of the interaction of contextual and perceptual factors on object identification was proposed.

The model proposed can be considered a more specified version of the intralevel priming hypothesis. It assumes a network of pictogens, with the pictogens linked according to semantic relatedness. Activation is assumed to travel along the links, allowing identified objects to prime the pictogens of related objects. Each pictogen is also assumed to acquire activation from perceptual sources, with the rate at which this activation accrues determined by the
quality of the perceptual input. Finally, it is assumed that there is a limit to the amount of activation which can accrue before the target object is actually fixated. Using this simple model, a qualitative fit to the main trends in the data of both experiments was possible.

The question which now remains is whether the intralevel priming model can account for context effects when an actual scene is employed as context. In opposition to the schema model, the intralevel priming model would predict that the last one or two objects fixated, rather than the scene as a whole, should be the predominant cause of the context effects found in scenes. Since coherent scenes tend to contain many related objects, the effect of overall scene context found in previous studies may have been produced by intralevel priming. What is needed is a direct comparison of the schema and intralevel priming models.

One way to test the intralevel priming model further would be to examine the effects of related objects on object identification in non-scene displays of more than two objects. Since the schema model posits that a coherent scene, with all objects in their correct spatial positions, is necessary in order to access an appropriate schema, it has no way of accounting for context effects produced in such displays. In fact, Biederman (personal
communication) believes that if displays like this are presented tachistoscopically, a "pop out" effect will occur such that an object unrelated to the other objects in the display will be identified more easily than an object related to the others. In contrast, the intralevel priming model would predict that an object related to the others will still be identified more rapidly due to spreading activation among the related objects.

Several studies are currently underway which attempt to examine these opposing predictions. In one set of studies being conducted by Peter DeGraef, groups of related and unrelated objects are displayed tachistoscopically, and the accuracy of identifying a target object in the parafovea is examined. In these studies, DeGraef is also manipulating the relatedness of a centrally fixated object to the target object in order to determine the relative benefit derived from foveal and parafoveal related primes. In a second set of experiments, I am using non-scene displays of related and unrelated objects to examine their effects on eye movement variables such as saccade length, fixation probability, and fixation duration. If non-scene displays produce context effects similar to those found with normal scenes, then it will be clear that a schema model need not be invoked in order to explain these effects.

Perhaps the most direct contrast of the schema and
intralevel priming models would be to embed two objects related to each other in a scene context in which they did not belong. For example, an octopus and a shark could be placed in an outdoor farm scene, where they would be anomalous. Previous research has shown that if the octopus alone were placed in such a scene, it would be fixated longer than if it were in an underwater scene. However, if a shark had been fixated prior to fixation on the octopus, would this result still be found? If so, the result would suggest that the schema model is indeed correct. However, if not, then this would suggest that the intralevel priming model is correct. Something in between would suggest that perhaps both play a role in scene processing.

Throughout this thesis I have presented the schema model and the intralevel priming model as mutually exclusive and opposing explanations of the effects of context on object identification. The presentation of these models as dichotomous stresses the differences between them and motivates research into the effects of context that otherwise would not be considered. Certainly the models do use entirely different types of processes in order to account for the context effects. However, the experiments presented in this thesis were meant to show that the intralevel priming model provided at least a motivated alternative
to the schema model, not to provide evidence that the schema model is wrong. Such evidence of the inadequacy of the schema model, if it is to be forthcoming at all, will have to await future research.

A final point is the possibility that both the schema model and the intralevel priming model are correct. It may be, in fact, that both the overall scene and the relations between the individual objects in the scene influence object identification. This possibility could easily be integrated into the model presented in Chapter 3 by allowing input to the pictogen from the scene level. However, it should be stressed that there is currently no reason to propose such a change to the model.

In sum, it could perhaps be said that the main contribution of this thesis will be to force the advocates of the schema position to prove that schemas are necessary in the face of an alternative account.
REFERENCES


Perception & Psychophysics, 25, 221-224.


Rayner, K., McConkie, G. W. (1976). What guides a


