Object Identification in Context: The Visual Processing of Natural Scenes

JOHN M. HENDERSON *University of Alberta*

Abstract When we view a natural visual scene, we seem able to determine effortlessly the scene's semantic category, constituent objects, and spatial relations. How do we accomplish this visual-cognitive feat? The commonly held explanation is known as the schema hypothesis, according to which a visual scene is rapidly identified as a member of a semantic category, and predictions generated from the scene category are then used to aid subsequent object identification. In this paper I will first outline and offer a critique of the evidence that has been taken to support the schema hypothesis. I will then offer an alternative framework for understanding scene processing, which I will call the local-processing hypothesis. This hypothesis assumes a modular, informationally-encapsulated architecture, and explicitly includes the role of covert visual attention in scene processing.

Résumé Lorsque nous voyons une scène visuelle naturelle, il semble que nous puissions déterminer sans effort la catégorie sémantique de cette scène, les objets qui la composent et les relations spatiales. Comment accomplissons-nous cet exploit cognitif-visuel? L'explication habituellement donnée se résume par l'hypothèse du schéma, selon laquelle une scène visuelle est rapidement associée à une catégorie sémantique et les prédictions découlant de cette catégorie permettent d'identifier l'objet par la suite. Dans ce document, j'exposerai d'abord les éléments de preuve qui militent en faveur de cette hypothèse et j'en ferai une critique. Je proposerai ensuite un autre cadre permettant de comprendre le traitement des scènes, soit l'hypothèse du traitement local. Cette hypothèse repose sur l'existence d'une architecture modulaire contenue dans l'information; elle inclut explicitement le rôle de l'attention visuelle voilée dans le traitement des scènes.

Review and Critique of Scene Processing Research

Research in scene processing has tended to focus on two main issues. First, how is it that a scene is identified as an instantiation or exemplar of a general class of scenes? For example, how is it that when I look out my office window, I know I am looking at a city scene (downtown Edmonton), rather

Canadian Journal of Psychology, 1992, 46:3, 319-341

than at an office scene, a bedroom scene, or a hockey scene? Does the visual system infer the type of scene from the identities of one or several diagnostic objects alone (Antes, Mann, & Penland, 1981; Friedman, 1979), or from a combination of those objects and their spatial relations (De Graef, Christiaens, & d'Ydewalle, 1990)? Or does the system instead identify the scene from scene-level features without first identifying particular objects (Biederman, 1981; Biederman, Glass, & Stacy, 1973)?

Second, what is the influence of the contextual constraint provided by a predictive scene on the identification of its constituent objects? For example, does the context serve to facilitate identification procedures for objects consistent with that scene? Concretely, can a cow be identified more accurately and/or more quickly if it is viewed in a farm scene rather than in a kitchen scene? A positive answer to this question would indicate that the object identification system can be influenced by contextual information represented at higher levels of analysis, and would imply that the object identification system is non-modular with respect to the scene level. In contrast, a negative answer would be consistent with the position that object identification is a modular system, informationally encapsulated from the scene level (Fodor, 1983).

The Schema Hypothesis as the Modal Model

The predominant view of the relation between object and scene identification can be summarized as the *schema hypothesis*. While there are several variations on the schema theme (e.g., Antes & Penland, 1981; Biederman, 1981; Friedman, 1979; Loftus & Mackworth, 1978), several commonalities define the hypothesis. According to the schema hypothesis, a memory representation of a prototypical scene is quickly activated during scene viewing, and is used to develop expectations about likely objects. These expectations then influence the object identification processes. For example, on the schema hypothesis, the identification of a cow in a farm scene involves (1) quickly recognizing that the scene is an exemplar of the category "farm scene", (2) accessing from memory the schema for a farm scene, (3) using the information stored with the schema to generate "cow" and other object candidates likely to be found in a farm scene (and possibly their canonical spatial relationships), and (4) using the knowledge that a cow is likely in such a scene to aid object identification processes when the cow is encountered.

While the above description suggests a serial model, this need not be a central assumption of the schema hypothesis. Identification of the farm scene and the individual cow could take place in parallel, but with mutual facilitation. This type of model can be conceptualized in terms of a connectionist architecture similar to the McClelland and Rumelhart (1981) interactive activation model of word recognition (e.g., Metzger & Antes, 1983), with objects corresponding to letters and scenes corresponding to words. On this view, activation of the "scene node" would activate potential object candidates, as the schema hypothesis requires. However, such an interactive activation account (and, indeed, any schema account) would have to surmount at least two problems. First, if the spatial relationships between objects are important in activating a particular set of object and scene nodes (as suggested by Biederman, Mezzanotte, & Rabinowitz, 1982), then location information would have to be explicitly encoded. This is, of course, also true of word recognition, where the spatial positions of letters are important in determining the particular word. However, to date the problem of coding spatial relations in connectionist models has proved difficult, and in models of word recognition, serial position has generally been added in the input representation rather than in the stored representation (see, e.g., Seidenberg & McClelland, 1990; see also Pinker & Prince, 1988, for a critique of this approach). In scene processing, the problem is compounded by the fact that there are more degrees of freedom in where objects can be located within a scene in comparison to letters within a word. The word CAT, for example, can be present if and only if the letter in the first position is a C. In an office scene, on the other hand, a phone can appear on a desk, on a table, or in the case of my office, on a modem atop a computer. Thus, the constraints on spatial relations in scenes are far less rigid than in words. Second, the presence of particular objects within a scene is probabilistic rather than certain, unlike letters in words. Thus, an office need not have a telephone at all, but the word CAT must contain a specific set of three letters. Even highly diagnostic objects within scenes need not appear with certainty. For example, a kitchen without a refrigerator would still be a kitchen, and would be recognizable as such. The word CAT without the C, on the other hand, would be another word. Thus, the constraints on constituents in a scene are also less rigid than in words.

In summary, the schema hypothesis can be decomposed into two main assumptions. First, it is assumed that the "gist" or category of a scene is determined very early during the processing of a new scene. Second, it is assumed that a determination of the scene category activates a schema that provides top-down information to facilitate identification of individual objects within that scene. While these two assumptions are logically separable, they have tended to co-occur within the schema hypothesis.

EXPERIMENTAL PARADIGMS

Two paradigms have traditionally been used to explore object identification in scenes, one involving object detection in tachistoscopic displays, and the other involving eye movement recording during free scene viewing. (The stimuli employed in these studies are line drawings of scenes, and the expectation is that the findings will generalize to natural scene viewing.) Results from both of these paradigms have been taken to support the schema hypothesis. In this section, I will present a brief summary of these paradigms and some of the problems with them. Further discussion of the methodologies used to explore scene processing can be found in Boyce and Pollatsek (1992) and Kroll (1992).

In the object detection paradigm pioneered by Biederman and his colleagues (e.g., Biederman, 1972; Biederman et al., 1973; Biederman et al., 1982; see also Boyce, Pollatsek, & Rayner, 1989; Murphy & Wisniewski, 1989), a scene is presented in a brief, masked display, and the subject is asked to determine whether a target object was present or absent at a cued location within the scene. The dependent measure is the probability of correctly detecting (or rejecting) the cued object as a function of the coherence of the scene (Biederman, 1972; Biederman et al., 1973), or as a function of the degree of congruence of the object with the scene (Biederman et al., 1982; Boyce et al., 1989). In the eye movement paradigm, subjects' eye movements are recorded while they visually explore a scene (Antes & Penland, 1981; De Graef et al., 1990; Friedman, 1979; Loftus & Mackworth, 1978). The eye movement record consists of rapid eye movements (saccades) and brief pauses (fixations) (Rayner, 1978; Yarbus, 1967). Because visual information is taken in only during the fixations, the amount of time spent fixating an object can be used as a measure of underlying processing difficulty (Rayner, 1978). The dependent measure in this paradigm, therefore, has been the fixation duration on an object as a function of the congruence of the object with the scene (Antes & Penland, 1981; De Graef et al., 1990; Friedman, 1979; Loftus & Mackworth, 1978).

The results of the object detection and eye movement paradigms have been taken to support both aspects of the schema hypothesis. First, results from these paradigms appear to support the view that the semantic category of a scene can be apprehended very rapidly. In the object detection task, effects of the scene context on detection accuracy are observed in brief (e.g., 150 ms) tachistoscopic displays, suggesting that the meaning of the scene (its semantic category) is apprehended during this brief exposure. In the eye movement paradigm, it has been shown that subjects move their eyes rapidly to an "informative" region of a scene (Antes, 1974; Mackworth & Morandi, 1967), sometimes after a single eye fixation (Antes, 1974; Loftus & Mackworth, 1978). Again, the suggestion is that because informativeness is defined in terms of the overall meaning of the scene, the meaning must have been quickly understood.

Second, results from both paradigms appear to support the view that object identification is facilitated when an appropriate scene schema has been activated. In the object detection paradigm, performance is facilitated when the target object appears in a coherent versus a jumbled scene (Biederman, 1972; Biederman et al., 1973; Biederman, Rabinowitz, Glass, & Stacy, 1974), when an object is probable versus improbable in the scene (Biederman et al., 1982; Boyce et al., 1989; Murphy & Wisniewski, 1989), and when the spatial relationship of the object to the scene is normal versus abnormal (Biederman et al., 1982; Boyce et al., 1989). In the eye movement paradigm, fixation duration on an object is shorter when the object is probable versus improbable in the scene (Antes & Penland, 1981; De Graef et al., 1990; Friedman, 1979; Loftus & Mackworth, 1978), and when the object is in a normal versus abnormal spatial position within the scene (De Graef et al., 1990). For the most part, the conclusion from both paradigms has been that the schema activated by the scene facilitates the identification of objects consistent with that schema.

Criticisms of the object detection paradigm

In the original object detection paradigm, subjects were presented with photographed scenes and were asked to 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 (Biederman, 1972; Biederman et al., 1973; Biederman et al., 1974). 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). The main result was that objects were detected more accurately in an intact versus a jumbled scene. The conclusion was that objects are more easily identified when a schema for a scene can be activated than when it cannot.

The finding that objects are detected more easily in intact versus jumbled scenes does not necessarily implicate the influence of scene-level context. In addition to disrupting the subject's ability to ascertain the semantic category of a scene, jumbling also adds, deletes, changes, and generally disrupts scene 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 visual noise. Under such circumstances, processes operative in object identification would be expected to suffer even under conditions in which the subject knows where to look and what to look for (as found by 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.

In order to circumvent the problems with the jumbling manipulation, recent studies using the object detection paradigm have examined scene context by manipulating the relationship of the target object to an otherwise intact scene (Biederman et al., 1982; Boyce et al., 1989; Murphy & Wisniewski, 1989). This revised object detection paradigm generally proceeds as follows: First, the subject is given the name of a target object; second, a scene is displayed briefly; third, a pattern mask is displayed containing a spatial cue; finally, the subject responds whether the target object occupied a location indicated by the spatial cue. In these experiments, "semantic" and "syntactic" relations are manipulated in a scene (Biederman et al., 1982). Semantic relations are the *probability* of an object occurring within the scene, the *position* of an object within the scene, and the *size* of an object given its location within the scene. Syntactic relations are *support* for an object which should rest on something, and *interposition* or occlusion of an object when it occurs behind another. For example, in a living room scene, a floating couch would violate the syntactic relation of support, while a fire hydrant atop a mail box would violate the semantic relation of position. The effects of manipulating these relations have been examined in object detection and violation detection (Was the object violating a relation?) tasks. Based on the findings that both semantic and syntactic relation violations equally affected detection performance in these tasks, and further that semantic relation violations added to the disruption in task performance caused by syntactic 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). The implication is that semantic relations are computed simultaneously with a physical parsing of the scene and before object identification has been completed. Thus, these results are taken to support both components of the schema hypothesis. First, the meaning of the scene is apprehended prior to object identification, and indeed prior to a complete physical parse of the objects within the scene. Second, activation of the scene schema makes available information about likely objects and object relations that can be used to facilitate object identification.

Although the manipulation of object relations in intact scenes addresses the main criticism of the jumbling manipulation, these experiments are still difficult to interpret. First, the distinction between semantic and syntactic violations is not a principled one. While there are undoubtedly some purely syntactic (i.e., structural) violations that could occur in scenes, such as those that lead to "impossible figures" as seen, for example, in Escher prints, it is likely that the syntactic violations explored to date are semantically based. (In fact, it is interesting to note that the structural violations in such impossible figures are often not immediately noticed by the viewer. It seems that such structural relations are not computed simultaneously over the entire image.) Whether an object requires visible support (for example, birds, balloons, airplanes, ceiling lights, and wall hangings do not), and whether an object should allow interposition (i.e., should be transparent, such as objects made of glass), depends on the identity of the object. Therefore, the finding that violations of support affect the same stages of processing as violations of position, size, or probability may be taken to show not that semantic representations of a scene are available as quickly as a physical parse of the scene, but instead to show that one type of semantic relation is available at about the same time as another. Further, even if one accepts the claim that interposition is determined at a strictly physical level, it is worth noting that in general, across all of the experiments discussed in Biederman et al. (1982), interposition violations have little if any effect at all (see also De Graef et al., 1990). Second, as Marcel (1983) demonstrated, pattern masks do not affect all representational levels in the same way. In particular, pattern masks seem to affect the availability of physical representations of visual input while leaving semantic representations relatively unaffected. Therefore, even if there were a distinction between semantic and syntactic violations, the use of a pattern mask following scene presentation in experiments designed to contrast the two would be expected to increase the effects of semantic relations compared with physical relations.

A second general criticism of the object detection paradigm is that object detection is not identical to object identification (e.g., Pashler & Badgio, 1987; Rabbitt, 1978). In object detection, because the name of the object is presented prior to presentation of the scene, subjects may generate some feature or subset of features to match to the visual stimulus in order to respond. This differs from the more natural case of identification, where a representation of the input must be constructed and matched with pre-stored memory representations of objects. This criticism could be answered if the object detection paradigm were changed such that the object name was provided following presentation of the scene. This type of study has to my knowledge been conducted once (Biederman et al., 1974), and, supportive of the schema hypothesis, the time of presentation of the target name (before or following presentation of the scene) did not interact with the context manipulation. Unfortunately, that study employed the jumbling manipulation. Thus, it is still unclear whether the same results would be found in the intact-scene version of the object detection task if the name of the target object were presented following the scene.

A third criticism of the object detection paradigm is that it is unclear what level of processing this paradigm reflects. Object identification can be conceptualized as one in a series of processing stages that are required to construct and encode in memory an integrated scene representation. In order to isolate the object identification stage from later stages in the processing sequence, it is necessary to choose an appropriate measure of object identification. Any task that purports to reflect identification processes must be able to separate identification from later, post-identification processes such as memory consolidation and response generation. This issue has long been recognized in the word recognition literature (e.g., Forster, 1979; Seidenberg, Waters, Sanders, & Langer, 1984), but has been less discussed in the scene processing literature. Preferably, one would use an on-line measure of performance, where an on-line measure can be defined as one that reflects a representation as it is being constructed. Ideally, such a measure would be unaffected by processes and representations occurring after the object identification stage.

Because identification and post-identification processes have not been separated in the object detection paradigm, scene context may be affecting at least two other processing levels. First, because the object detection paradigm employs a two-response forced-choice decision component, subjects may use sophisticated guessing strategies to decide whether the pre-cued target object occurred within the scene (see, e.g., De Graef et al., 1990, for explication). Thus, context may exert its influence on a post-identification, response generation stage of processing. This possibility was recognized early in the word recognition literature, and steps were taken to guard against such strategies in explorations of the word-superiority effect (Reicher, 1969; Wheeler, 1970). Similar steps, such as presenting the subject with two equally likely target object candidates following presentation of a scene, could help to overcome this problem in scene processing as well. In fact, a recent study by Masson (1991) using this type of forced-choice paradigm suggests that scene context does not influence visual analysis, but does influence the criterion amount of information required to decide that a particular object is present.

Second, as argued by Henderson, Pollatsek, and Rayner (1987), scenes may activate memory schemata (rather than perceptual schemata) into which congruent objects are more easily integrated. Objects which can be easily integrated into such a representation may be facilitated at the time of response in the object detection paradigm. On this view, objects are *identified* equally well in tachistoscopic presentations regardless of whether or not they are congruent with the scene (e.g., whether or not they are violating semantic or syntactic constraints). Instead of feeding information top-down to identification routines, the schema would affect the availability of information at the time of response, either because objects that did not fit easily into the memory schema would never be included in the memory representation of the scene, or because they would be included but would be more difficult to retrieve from the memory representation during response generation. The memory schema explanation differs from the perceptual schema explanation in that the former does not postulate any effects of context on object *identification* processes, only on memory integration and/or memory retrieval processes following identification.

Support for the memory explanation derives from the work of Potter (1975, 1976) and Intraub (1979, 1980, 1981), who have shown that objects may be very quickly identified but not remembered if masked by a following visual stimulus. In addition, the memory explanation predicts facilitation for an object that can easily be integrated into a memory representation of a scene regardless of whether or not the object is predictable in the scene. The standard perceptual schema hypothesis, on the other hand, predicts facilitation only for objects that are predictable from the scene schema. The results of the object detection study by Boyce et al. (1989) are consistent with the memory explanation but not with the schema hypothesis: Objects that were consistent but not predictable in a scene were as facilitated as objects that were both consistent and predictable.

Criticisms of the eye movement paradigm

Two issues must be dealt with in order to use the eye movement paradigm to explore scene processing. First, in order to use any fixation time measure as an indication of object identification time, it must be demonstrated that the measure reflects identification but not other, post-identification processes. It is likely that global measures of fixation time, such as the total time spent on an object during the course of scene viewing, and the gaze duration on an object (the time of all initial fixations on an object prior to leaving that object for the first time, Just & Carpenter, 1980) reflect post-identification processes. In reading, for example, gaze duration has been found to reflect syntactic parsing (Ferreira & Henderson, 1990; Frazier & Rayner, 1982) and semantic integration (Ehrlich & Rayner, 1983; Inhoff, 1984). Thus, it is likely that gaze duration in scene processing reflects other processes beyond object identification. Unfortunately, most of the early eye movement studies used the gaze duration measure (Antes, 1974; Antes & Penland, 1981; Friedman, 1979; Loftus & Mackworth, 1978), although several of these authors called their measure "first fixation duration" (Antes & Penland, 1981; Friedman, 1979). Therefore, demonstrations of scene context effects in the eye movement paradigm probably indicate that context influences processes such as the overall understanding of the object's role within the scene (e.g., what was that cow doing in a kitchen?) rather than the time to identify the object. The preferred fixation measure is the true first fixation duration (De Graef et al., 1990; Henderson et al., 1989), or the duration of time from the initial landing of the eyes on an object until the eyes move to any other location, including another location on the object. This measure is necessarily shorter than other measures, and therefore is more likely to reflect early processes alone. Unfortunately, however, there is as yet no good evidence that first fixation duration does not also sometimes reflect later post-identification processes.

Second, the basic premise of the eye movement paradigm is that the results will reflect normally occurring visual-cognitive processes because subjects can view scenes in a natural manner. However, unlike reading, where the overall task is arguably transparent, subjects must be given an orienting task when they view a scene. In the majority of studies conducted to date, subjects have been told that they are to examine the scenes in preparation for a subsequent memory task (e.g., Antes & Penland, 1981; Friedman, 1979; Loftus & Mackworth, 1978; but see De Graef et al., 1990). Unfortunately, viewing behaviour and eye movement patterns change as a function of the viewing task given to the subject (Yarbus, 1967). It is unclear whether subjects in these studies look for a longer period of time at objects that are incongruent rather than congruent with a scene because the objects take longer to identify, or because they take longer to integrate into a sustainable memory representation (see Boyce & Pollatsek, 1992, for similar arguments).

One way to address the orienting task issue would be to give subjects a task that did not force the creating of a coherent memory representation of the scene, and look for similar scene context effects on fixation time across tasks. Recently, such a study was conducted by De Graef et al. (1990), in which the orienting task required subjects to count the number of non-objects (Kroll & Potter, 1984) in a scene. Interestingly, while De Graef et al. found effects of probability, position, and support on first fixation duration, these effects only appeared on objects that were fixated relatively late in scene viewing, following on average the first eight fixations on the scene. These results thus provide evidence against the view that the context provided by a scene necessarily produces immediate effects on object processing. It is not clear whether the effect of the scene context came into play late in the De Graef et al. study because the meaning of the scene was apprehended late, or because the meaning of the scene was apprehended early but did not exert an influence on object processing until later. In any case, the De Graef et al. results suggest that further work will be needed to determine how large a role the orienting task plays, and which type of orienting task will best lead to insights into the initial identification of the scene's meaning and the identification of the objects within the scene.

GENERAL CRITICISMS OF THE SCHEMA HYPOTHESIS

Aside from the specific problems for the schema hypothesis 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. First, should contextual influences be considered the default theoretical position? For many years, the common wisdom in computer vision was that all knowledge sources need to be consulted in a highly interactive way in order to identify an object. This type of view was accepted into cognitive psychology, and is one of the traditions from which schema theory in visual cognition derives. However, more recently it has become clear that models of this type have severe problems, and that a good deal of progress can be made without assuming strongly knowledge-driven processing (Marr, 1982). If we choose 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. If, after a careful investigation, 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 discovered.

Second, because the schema hypothesis is 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. 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, 1987; Intraub, 1979, 1980, 1981; Potter, 1975, 1976). In fact, when misidentification does occur (for example, when a cardboard box on the road is taken as an animal under the degraded visual conditions of a dark night), it is usually quite striking. The fact that misidentification is striking when it occurs suggests that it doesn't occur very often. The question is whether it is reasonable to posit that the human visual system is as easily fooled as schema theories seem to imply.

A third question that the schema hypothesis 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, 1985). Friedman (1979) has been most explicit in her description of how a schema facilitates object recognition. On her view, resource-free feature matching takes place for objects predicted by the schema, but resource-intensive feature analysis takes place for objects that are not predicted. However, it is not at all clear that other schema theorists accept her view. Certainly the predicted effects of schema activation will be different depending on the manner in which the knowledge contained in a schema is used.

A final concern for any theory of perception that allows general semantic knowledge to influence perceptual processes is the "frame problem". The problem is that of determining which aspects of all available world knowledge are relevant to a particular situation. For example, Biederman et al. (1982) propose that information about the normal position of an object in relation to its visual context is contained in a schema. 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).

SUMMARY

The results derived from both the object detection and eye movement paradigms have led most scene theorists to the view that scene processing involves the rapid activation of a schema that is then used to facilitate subsequent object identification. However, there are sufficient problems with both paradigms to warrant caution at the empirical level. At the least, the data severely underdetermines the schema hypothesis. Further, there are theoretical reasons for remaining sceptical about the need for postulating the type of top-down information flow suggested by the schema hypothesis. In the remainder of this article, I will outline an alternative theoretical framework from which to view the scene processing literature.

The Local Processing Hypothesis

In this section, I want to propose an alternative to the schema hypothesis. This alternative, which I will call the *local processing hypothesis*, is founded upon two underlying assumptions. The first assumption is that the object identification system is informationally encapsulated (Fodor, 1983) from the system that determines the overall category of the scene. The essential notion is that knowing the semantic category of the scene can have no influence on processes or representations at the level of object identification. Instead, context effects in scenes can be considered local in that they are limited to interactions at the level of object representations. The types of information which an informationally encapsulated module can use are representations which must be computed prior to the operation of that module and which serve as its input (e.g., features, contours, or parts in the case of object identification), and information which is necessarily contained within the module (e.g., stored object representations and the structure which organizes those representations and makes them retrievable). On this view, then, the information that is available to the object identification stage is perceptual information from lower levels of processing and information about object representations (intralevel information), but *not* higher level information. The local processing hypothesis represents an explicit rejection of the assumption that predictions about likely objects and scene-specific spatial relations generated from knowledge of the scene category can affect object identification processes. While the local-processing hypothesis does not deny that these predictions can be generated, it does deny that the predictions are available to the object identification process.

There are several reasons for assuming that the object identification system is informationally encapsulated. First, as Fodor states, "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). Second, as discussed above, it seems more likely that we will discover structure in the cognitive system if we explicitly look for it than if we assume an undifferentiated, interactive system (Forster, 1979; Tanenhaus, Dell, & Carlson, 1987). Third, much work in psycholinguistics suggests that aspects of such processes as visual word identification (Duffy, Henderson, & Morris, 1989; Foss, 1988; Kintsch & Mross, 1985; Masson, 1988; Seidenberg, 1985; Stanovich, 1991), auditory word recognition (Connine, 1987), lexical meaning retrieval (Seidenberg, Tanenhaus, Leiman, & Bienkowski, 1982; Swinney, 1979), and syntactic parsing (Ferreira & Clifton, 1986; Ferreira & Henderson, 1990; Frazier & Rayner, 1982) are informationally encapsulated, suggesting that other cognitive processes may be as well.

Recall that there were two assumptions in the schema hypothesis, rapid scene categorization and top-down facilitation of object identification. While information encapsulation is in direct opposition to top-down processing, it does not conflict with rapid categorization. Information encapsulation implies only that scene knowledge should not influence object identification. The point, according to the information encapsulation assumption, is that even if the scene category does become available very early, predictions about the identities and locations of objects in the scene generated from the scene's category should not be able to influence identification processing operating on objects in the scene.

The second assumption of the local processing hypothesis is based upon the finding that during stimulus identification, attention is allocated to a limited region of the visual field, both in tachistoscopic studies (e.g., Treisman, 1988; Treisman & Gelade, 1980) and prior to saccadic eye movements (Henderson & Ferreira, 1990; Henderson, Pollatsek, & Rayner, 1989; Rayner, McConkie, & Ehrlich, 1978; see Henderson, 1992a, for a review). These findings have led many theorists in visual cognition to believe that the amount of a visual scene that can be analyzed semantically at any one time is spatially limited (e.g., Marr, 1982; Pylyshyn, 1988; Treisman, 1988; Ullman, 1984). More specifically, when objects have to be identified and/or selected for further processing, it appears that the system must allocate limited attentional resources to as few as one object at a time. Thus, scene processing according to this assumption can be considered local in that interactions among scene elements are likely to take place within spatially limited attended regions.

ATTENTION, CONTEXT EFFECTS, AND INTRALEVEL PRIMING

According to the local-processing hypothesis, context effects are due to interactions among object-level representations computed within limited, attended regions. One example of this type of interaction is *intralevel priming* (Henderson et al., 1987), wherein representations of objects that have recently been identified at an attended location prime the representations of semantically or associatively related objects. According to the intralevel priming view of object identification, probability (congruity) effects in scenes occur as a result of the organization of the representations within the object recognition module. Simple associative, semantic, or episodic 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 consulted in computing an output.[1]

Many studies have demonstrated that object identification can be primed by prior presentation of a related object (e.g., Carr, McCauley, Sperber, & Parmelee, 1982; Huttenlocher & Kubicek, 1983; Kroll & Potter, 1984). Building on these results, Henderson et al. (1987) showed in both naming-latency and fixation-time paradigms that priming can occur between objects viewed over successive eye fixations (see also Henderson, 1992b). Henderson et al. (1987) argued from this result that the probability (congruency) context effects observed in scene experiments could be explained by this type of object-to-object priming. I also recently found a similar priming effect when the prime object flanked the location of the target object prior to an eye movement to that location (Henderson, 1992c), particularly when the target object was difficult to identify. Therefore, it may be that in scenes one extrafoveal object can be primed by another related extrafoveal object if the two are spatially close and within the attended region of the scene.

According to the intralevel priming explanation, objects are processed faster when they are viewed in a scene context because they are primed by previously or simultaneously attended objects, not because of the overall meaning of the scene itself. In the case of the eye movement studies, a fixated object would be primed by the object viewed on the preceding fixation, as well as objects located near the fixated object. In the case of the object detection paradigm, attention may be allocated first to the centre of the scene (where fixation is held), next to one or more regions of the scene prior to scene termination, and finally to the cued region of the representation of the scene that remains active following presentation of the mask. (The cue itself may prevent the cued area from being masked as effectively as the rest of the

1 It is often claimed that massively parallel systems make this form of argument obsolete, because all sources of information can be considered simultaneously. However, it has recently been suggested that even massively parallel models may have to be relatively modular between levels of processing (e.g., Norris, 1990; Tanenhaus et al., 1987).

scene.) In this case, priming could occur from related objects at fixation, at other attended locations, and from objects surrounding the cued object.

Objections to Intralevel Priming

Three objections have been raised against intralevel priming as a complete explanation of the effects of a scene context on object identification. First, it has been suggested that objects in scenes are episodically rather than semantically related (Boyce et al., 1989). In other words, a barn and a cow are better thought of as related by virtue of co-occurrence rather than by virtue of sharing semantic features. Thus, demonstrations of semantic priming across eye fixations may say little about context effects in scenes because few objects are likely to be semantically related. There are two responses to this criticism. First, it is not unlikely that semantically related objects would tend to co-occur in the same scene. For example, consider the object pairs used in Henderson et al. (1987) and Henderson (1992c) experiments. In each case, it is easy to specify a scene that would contain each pair of objects (e.g., a cow and horse in a farm scene, a shirt and pants in a laundry scene, a comb and brush in a bathroom scene, a hammer and saw in a toolbench scene, etc.). Second, episodic relations appear to be prestored in memory along with semantic relations. In support of this view, De Graef (1992) has explicitly examined foveal priming across eye movements in non-scene arrays for objects that are episodically but not semantically related. Consistent with Henderson et al. (1987), De Graef (1992) found that fixation durations on target objects were reduced following fixations on episodically related objects.

A second objection to the intralevel priming account is the existence of context effects that cannot be explained in terms of simple priming (De Graef et al., 1990). For example, as described above, Biederman et al. (1982) provided evidence for effects of three "semantic" context effects (probability, position, and size) and two "syntactic" context effects (support and interposition). If the observed effects truly indicate that other relations beyond probability affect object identification, then intralevel priming could not offer a complete account of context effects in scenes. In response, however, it is not clear how robust the various violation effects are. Virtually all studies employing both the object detection and eye movement paradigms have provided evidence for probability violations. On the other hand, only two published studies that I know of have examined effects of the other violations. One of these studies employed the object detection paradigm (Biederman et al., 1982) and the other employed the eye movement paradigm (De Graef et al., 1990). A careful examination of these studies suggests that the evidence for context effects beyond simple probability effects is marginal. In the Biederman et al. (1982) study, once camouflage (due to the number of nearby contours) across context conditions was controlled, effects of size, position, and probability were clearly found. There was no effect of interposition, and there was an apparent speed-accuracy tradeoff for support. In the De Graef et al. (1990) study, the effects of probability, position, and support were reliable. The only effect that replicates over paradigms (in addition to probability, which has been demonstrated many times), is the effect of position. Thus, the evidence for effects of other types of relations, particularly syntactic relations, is weak.

A final objection to the intralevel priming explanation of scene context effects is the claim that context effects do not appear in non-scene arrays of objects (Biederman, Blickle, Teitelbaum, & Klatsky, 1988; Boyce et al., 1989). For example, Boyce et al. (1989) found that object detection was better in a full scene compared with a display consisting of the same objects but without the scene background. If object identification alone were able to lead to the scene context effect, then similar effects should have been observed in the two conditions. However, a major difference between the scene and objects-only conditions was the lack of a foveal object in the objects-only condition. (There were similarly no foveal objects in the Biederman et al., 1988, study). In the scene condition, on the other hand, a large object forming part or all of the background was often present (e.g., a refrigerator for a refrigerator scene and a swimming pool for a swimming pool scene). Generally, when a foveal context object (or an extrafoveal but attended context object; Henderson, 1992c) is included in a non-scene array, context effects are observed (De Graef, 1992; Henderson, 1992b; Henderson et al., 1987). Therefore, it appears that a necessary prerequisite for intralevel priming to occur is that the priming object be attended.

ATTENTION, LOCAL PROCESSING, AND VIOLATION EFFECTS

While intralevel priming can account for probability effects on object identification, it cannot account for effects that depend on the spatial relations of the objects within a scene. As discussed above, the evidence for the view that spatial relations affect object processing is still not convincing. However, even if we assume that these other relations produce reliable effects, intralevel priming was not proposed in order to account for all violation effects. The specific point of the Henderson et al. (1987) study was that the majority of data supporting the schema hypothesis from both the object detection and eye movement paradigms had been based on probability manipulations, and that another account beyond schema activation for these effects was viable. A more general point is that what at first glance appear to be effects due to a scene-level representation may in fact be due to much more local computations between objects within relatively small attended regions.

Biederman et al. (1982) presented evidence that violations of semantic relations disrupt object detection as quickly as do syntactic violations. This finding was taken as evidence that the meaning of a scene (represented by a schema) became active very early during scene processing (prior to or simultaneously with identification of individual objects), and that information from the schema influenced subsequent object identification. An object-to-object priming account of probability (relatedness) effects is not inconsistent with the view that so-called semantic and syntactic relations are computed on an object simultaneously. (This argument again assumes that the distinction between semantic and syntactic relations can be maintained.) In fact, one might suppose that the object previously identified, or the flanker objects that are simultaneously attended, might bias the computation of physical relations. For example, computation of which contours belong with which object in a line drawing of a scene might partly depend on how various parses of the contours in a region interact with potential object interpretations, where some object interpretations are more active due to priming. What is at issue is not whether semantic relations have an early effect on object identification, but what the source of the semantic effect is. According to the schema hypothesis, semantic effects are due to activation of a scene-level representation that provides information about likely objects. According to the local-processing hypothesis (as instantiated, for example, in intralevel priming), semantic effects are due to interactions operating within the object identification module from other objects or scene properties that have recently been attended, but not to an understanding of the meaning of the scene as a whole.

ATTENTION AND RAPID COMPREHENSION OF A SCENE

As discussed above, the finding that a scene's semantic category can be accessed early in scene viewing does not violate the notion that the object identification system is informationally encapsulated. The finding does, however, raise the issue of attentional processing. If we assume that scene viewing requires a preattentive, parallel stage of analysis (to find visual primitives) followed by an attentive, sequential stage (to make available categorical information), then how can we account for the data suggesting that the semantic category of a scene is rapidly determined? There are at least three possibilities. First, it could be that the category is determined based on preattentive information alone. For example, perhaps a set of features such as a collection of principal axes of objects can specify general classes of scenes. If, however, as is the case of object identification, the spatial relations between such features are also important, so that the features need to be conjoined, then attentional processing should be required (e.g., Treisman & Gelade, 1980).

A second possibility is that some set of scene-level features, when combined via attentional processing, creates a scene-level description similar to an object description, but at a larger scale. This notion is related to Biederman's (1990) geon-based scene models, though it is not clear what role attention is to play on Biederman's view. A question on this second account is whether attention must be "spread over" the entire scene if a scene-level representation is going to be computed in a single glance. That is, perhaps it is only when a wide spread of attention conjoins scene-level features that a scene description is initially computed. In the object detection paradigm, the subject is induced to begin the trial with visual attention spread over as great a region as possible, because the spatial location cue can appear anywhere in the scene. If subjects allocate attention in this manner, then it may be that the tachistoscopic paradigm is biased toward providing evidence for rapid scene comprehension. This situation is clearly different from that of natural scene viewing, where attention is normally focused at the foveal location and the location about to be fixated (Henderson, 1992a). In this case, the scene-level description may emerge over time as objects and spatial relations are computed, as recently found by De Graef et al. (1990).

A third possible hypothesis regarding the role of visual attention in the rapid apprehension of the semantic category of a scene is the sequential, local processing explanation. Given that objects can be identified very rapidly (e.g., Potter, 1975, 1976; Intraub, 1979, 1980, 1981), and that attention can be shifted rapidly from one location to another without a concomitant eye movement (e.g., Remington & Pierce 1984; Sperling & Reeves, 1980), several objects in different spatial locations within the scene might be identified within a single fixation or single tachistoscopic glance. Given that the identification of one or several diagnostic objects can lead to semantic activation of a specific scene concept (Antes, Mann, & Penland, 1981; Friedman, 1979), it is possible that the rapid apprehension of a scene reflects rapid identification of one or several of these objects, particularly objects near the fovea.

Conclusion

In this paper, I examined some central issues in scene processing. I first presented a summary of the research on scene processing along with a description of the schema hypothesis. The schema hypothesis proposes that the semantic category of a scene is quickly apprehended, and that predictions about objects (and possibly spatial relations) from the scene category are then used to facilitate subsequent object identification. This hypothesis has been the primary theoretical construct used by researchers to account for the results of scene processing studies. However, I presented both empirical and theoretical reasons for remaining cautious about the schema hypothesis. Second, I presented an alternative explanation for the scene results, which I called the local-processing hypothesis. The central assumptions of this hypothesis are that object identification is an informationally encapsulated module, and that attention must be oriented to individual objects in order for those objects to be identified. Finally, I indicated how the results from previous scene processing studies can be accommodated by the local processing hypothesis.

The study of natural scene processing is still relatively new. Many questions remain unresolved, including how visual attention is allocated during the course of natural scene perception, what the implications of attentional orienting are for object identification and scene interpretation, how a scene concept is accessed, and whether and how contextual factors can influence the manner in which an object representation is constructed and matched to stored representations. The main point of the current paper was to suggest that the schema hypothesis provides but one possible answer to these questions. The data currently available are also consistent with a local processing alternative in which top-down influences are eliminated and context effects emerge from interactions among object identification routines operating on spatially limited regions of the scene.

Preparation of this article was supported by the Central Research Fund of the University of Alberta, and by the Natural Sciences and Engineering Research Council of Canada (OGP-41792). I thank Peter De Graef, Fernanda Ferreira, and Lise Paquet for commenting on earlier incarnations of this article. Please address correspondence to the author, who is now at the Department of Psychology, Michigan State University, East Lansing, MI, 48824, USA.

References

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

Antes, J.R., Mann, S.M., & Penland, J.G. (1981). *Local precedence in picture naming: The importance of obligatory objects.* Paper presented at the Annual Meeting of the Psychonomic Society.

Antes, J.R., & Penland, J.G. (1981). Picture context effect on eye movement patterns. In D.F. Fisher, R.A. Monty, & J.W. Senders (Eds.), *Eye movements: Cognition and visual perception.* Hillsdale, NJ: Erlbaum.

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

Biederman, I. (1981). On the semantics of a glance at a scene. In M. Kubovy & J.R. Pomerantz (Eds.), *Perceptual organization.* Hillsdale, NJ: Erlbaum.

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

Biederman, I. (1990). Higher-level vision. In D.N. Osherson, S.M. Kosslyn, & J.M. Hollerbach (Eds.), *Visual cognition and action.* Cambridge, MA: MIT Press.

Biederman, I., Blickle, T.W., Teitelbaum, R.C., & Klatsky, G.J. (1988). Object search in nonscene arrays. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 456-467.

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

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

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

Boyce, S.J., & Pollatsek, A. (1992). An exploration of the effects of scene context on object identification. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading*. New York: Springer-Verlag.

Boyce, S., Pollatsek, A., & Rayner, K. (1989). *Journal of Experimental Psychology: Human Perception and Performance, 15*, 546-556.

Carr, T.H., McCauley, C., Sperber, R.D., & Parmelee, C.M. (1982). Words, pictures, and priming: On semantic activation, conscious identification, and the automaticity of information processing. *Journal of Experimental Psychology: Human Perception and Performance, 8*, 757-777.

Connine, C.M. (1987). Constraints on interactive processes in auditory word recognition: The role of sentence context. *Journal of Memory and Language, 26*, 527-538.

De Graef, P. (1992). Scene context effects and models of real-world perception. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading*. New York: Springer-Verlag.

De Graef, P., Christiaens, D., & d'Ydewalle, G. (1990). Perceptual effects of scene context on object identification, *Psychological Research, 52*, 317-329.

Duffy, S.A., Henderson, J.M., & Morris, R.K. (1989). Semantic facilitation of lexical access during sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 791-801.

Ehrlich, K., & Rayner, K. (1983). Pronoun assignment and semantic integration during reading: Eye movements and immediacy of processing. *Journal of Verbal Learning and Verbal Behaviour, 22*, 75-87.

Ferreira, F., & Clifton, C.E. (1986). The independence of syntactic processing. *Journal of Memory and Language, 25*, 348-368.

Ferreira, F., & Henderson, J.M. (1990). The use of verb information in syntactic parsing: Evidence from eye movements and word-by-word self-paced reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 555-569.

Fodor, J.A. (1983). *Modularity of mind*. Cambridge, MA: MIT Press.

Forster, K.I. (1979). Levels of processing and the structure of the language processor. In W.E. Cooper and E. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett*. Hillsdale, NJ: Erlbaum.

Foss, D.J. (1988). Experimental psycholinguistics. *Annual Review of Psychology, 39*, 301-348.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology, 14*, 178-210.

Friedman, A. (1979). Framing pictures: The role of knowledge in automatized encoding and memory for gist. *Journal of Experimental Psychology: General, 108*, 316-355.

Henderson, J.M. (1992a). Visual attention and eye movement control during reading and scene perception. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading*. New York: Springer-Verlag.

Henderson, J.M. (1992b). Eye movement control during visual object processing: Effects of initial fixation position and semantic constraint. *Canadian Journal of Psychology*, in press.

Henderson, J.M. (1992c). Identifying objects across saccades: Effects of extrafoveal preview and flanker object context. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 521-530.

Henderson, J.M., & Ferreira, F. (1990). Effects of foveal processing difficulty on the perceptual span in reading: Implications for attention and eye movement control. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 417-429.

Henderson, J.M., Pollatsek, A., & Rayner, K. (1987). Effects of foveal priming and extrafoveal preview on object identification. *Journal of Experimental Psychology: Human Perception and Performance, 13*, 449-463.

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

Hochberg, J. (1978). *Perception*. Englewood Cliffs, NJ: Prentice-Hall.

Huttenlocher, J., & Kubicek, L.F. (1983). The source of relatedness effects on naming latency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*, 486-496.

Inhoff, A.W. (1984). Two stages of word processing during eye fixations in the reading of prose. *Journal of Verbal Learning and Verbal Behaviour, 23*, 612-624.

Intraub, H. (1979). Presentation rate and the representation of briefly glimpsed pictures in memory. *Journal of Experimental Psychology: Human Learning and Memory, 5*, 78-87.

Intraub, H. (1980). The role of implicit naming in pictorial encoding. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 1-12.

Intraub, H. (1981). Rapid conceptual identification of sequentially presented pictures. *Journal of Experimental Psychology: Human Perception and Performance, 7*, 604-610.

Julesz, B. (1971). *Foundations of Cyclopean Perception*. Chicago: University of Chicago Press.

Just, M., & Carpenter, P. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*, 329-354.

Kintsch, W., & Mross, E.F. (1985). Context effects in word identification. *Journal of Memory and Language, 24*, 336-349.

Kroll, J.F. (1992). Making a scene: The debate about context effects for scenes and sentences. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading*. New York: Springer-Verlag.

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

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

Mackworth, N.H., & Morandi, A.J. (1967). The gaze selects informative details within pictures. *Perception and Psychophysiology*, *2*, 547-552.

Marcel, A.J. (1983). Conscious and unconscious perception: Experiments on visual masking and word recognition. *Cognitive Psychology*, *15*, 197-237.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: W.H. Freeman & Company.

Masson, M.E.J. (1988). The interaction of sentence context and perceptual analysis in word identification. *Memory & Cognition*, *16*, 489-496.

Masson, M.E.J. (1991). *Constraints on the interaction between context and stimulus information*. Presented at the Thirteenth Annual Meeting of the Cognitive Science Society, Chicago, Illinois.

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

Metzger, R.L., & Antes, J.R. (1983). The nature of processing early in picture perception. *Psychological Research*, *45*, 267-274.

Murphy, G.L., & Wisniewski, E.J. (1989). Categorizing objects in isolation and in scenes: What a superordinate is good for. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 572-586.

Norris, D. (1990). How to build a connectionist idiot (savant). *Cognition*, *35*, 277-291.

Pashler, H., & Badgio, P.C. (1987). Attentional issues in the identification of alphanumeric characters. In M. Coltheart (Ed.), *Attention and Performance XII* (pp. 63-81). Hillsdale, NJ: Erlbaum.

Pinker, S. (1985). Visual cognition: An introduction. In S. Pinker (Ed.), *Visual cognition*. Cambridge, MA: MIT Press.

Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. In S. Pinker and J. Mehler (Eds.), *Connections and symbols* (pp. 73-193). Cambridge, MA: MIT Press.

Potter, M.C. (1975). Meaning in visual search. *Science*, *187*, 965-966.

Potter, M.C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory*, *5*, 509-522.

Pylyshyn, Z. (1988). Here and there in the visual field. In Z. Pylyshyn (Ed.), *Computational processes in human vision: An interdisciplinary perspective*.

Norwood, NJ: Ablex Publishing Company.

Rabbitt, P.M.A. (1978). Sorting, categorisation, and visual search. In E.C. Carterette and M.P. Friedman (Eds.), *Handbook of perception IX* (pp. 85-134). London: Academic Press.

Rayner, K. (1978). Eye movements in reading and information processing. *Psychological Bulletin, 85*, 618-660.

Rayner, K., McConkie, G.W., & Ehrlich, S. (1978). Eye movements and integrating information across fixations. *Journal of Experimental Psychology: Human Perception and Performance, 4*, 529-544.

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

Remington, R., & Pierce, L. (1984). Moving attention: Evidence for time-invariant shifts of visual selective attention. *Perception & Psychophysics, 35*, 393-399.

Seidenberg, M.S. (1985). Constraining models of word recognition. *Cognition, 20*, 169-190.

Seidenberg, M.S., & McClelland, J.L. (1990). A distributed, developmental model of word recognition and naming. *Psychological Review, 96*, 523-568.

Seidenberg, M.S., Tanenhaus, M.K., Leiman, J.L., & Bienkowski, M. (1982). Automatic access of the meanings of words in context: Some limitations of meaning-based processing. *Cognitive Psychology, 14*, 489-537.

Seidenberg, M.S., Waters, G.S., Sanders, M., & Langer, P. (1984). Pre- and postlexical loci of contextual effects on word recognition. *Memory and Cognition, 12*, 315-328.

Sperling, G., & Reeves, A. (1980). Measuring the reaction time of a shift of visual attention. In R. Nickerson (Ed.), *Attention and performance VIII* (pp. 347-360). Hillsdale, NJ: Erlbaum.

Stanovich, K.E. (1991). Word recognition: Changing perspectives. In R. Barr, M.L. Kamil, P. Mosenthal, and P.D. Pearson (Eds.), *Handbook of reading research (Volume 2)* (pp. 418-452). New York: Longman.

Swinney, D.A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behaviour, 18*, 645-659.

Tanenhaus, M.K., Dell, G.S., & Carlson G. (1987). Context effects in lexical processing: A connectionist approach to modularity. In J. Garfield (Ed.), *Modularity in knowledge representation and natural language understanding*. Cambridge, MA: MIT Press.

Treisman, A. (1988). Features and objects: The 14th Bartlett memorial lecture. *The Quarterly Journal of Experimental Psychology, 40*, 201-237.

Treisman, A., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology, 12*, 97-136.

Ullman, S. (1984). Visual routines. *Cognition, 18*, 97-159.

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

Yarbus, A.L. (1967). *Eye movements and vision.* New York: Plenum Press.