2011

The Role of Contextual Associations in the Selection of Objects

Noah Patrick Sulman
University of South Florida, noah.sulman@gmail.com

The Role of Contextual Associations in the Selection of Objects

by

Noah Sulman

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Psychology
College of Arts and Sciences
University of South Florida

Major Professor: Thomas Sanocki, Ph.D.
Kenneth Malmberg, Ph.D.
Toru Shimizu, Ph.D.
Sudeep Sarkar, Ph.D.
Dmitry Goldgof, Ph.D.

Date of Approval:
April 6, 2011

Keywords: attention, scene context, object recognition, attentional capture

Acknowledgements

I would like to thank Thomas Sanocki, whose patient encouragement over the past few years was invaluable. He knows when to criticize, when to help, and when to step back. Not a day passes when I don't use the critical thinking, technical, and communication skills I acquired while in this lab.

I must also thank my committee members, Kenneth Malmberg, Dmitry Goldgof, Sudeep Sarkar, and Toru Shimizu, for investing so much of their time and energy in ensuring my success. I would not have had an opportunity to develop my abilities as a scientist were it not for your continued thoughtful consideration. I would also like to thank the administration staff within the psychology department for constantly putting up with late and incomplete paperwork, last second changes, and many other unnecessary troubles. I certainly know my late forms made no one's job any easier.

Steve Fiske and Jen O'Brien, as fellow lab researchers, were an incredible resource, always willing to listen to me ramble and generous with insight and informed opinion.

Lastly, and most importantly, I would like to thank my families. Dick and Linda Sulman insisted I always ask questions. I haven't stopped. They probably regret it. Steve and Jill Bunker encouraged me to get as much of an education as possible. They may regret it as well (it sure took long enough). My wonderful wife Heather and daughter Lilian sustain me with their loving kindness. I wake up everyday to make sure they never regret it.

Table of Contents

List of Figures iv

Abstract v

Chapter 1: Visual Attention 1
- Introduction 1
- Human Visual Attention 4
- The Medium of Visual Attention 6
  - Spatial attention models 6
  - Perceptual organization based models 8
- Attention in Theoretical Accounts of Visual Cognition 13
  - Feature integration theory 13
  - Coherence and FINST theories 16
- Control of Visual Attention 20
  - Top-down and bottom-up factors in ACS 20
  - Attentional control and perceptual organization 24
  - Locus of attentional selection 32
- Conclusions 36

Chapter 2: Object-context Associations in Object Recognition 37
- Direct measures of contextual influences on object recognition 37
  - Indirect measures of contextual influence 44
- A model of object recognition using contextual information 49
- Contextual Associations in Visual Search 50
  - Contextual cueing 51
  - Models of contextual influences on scene search 52
- Attention and Temporal Limits in Perception 57
  - Attention and conceptual short-term memory 57
  - Rapid access to affective information 62
  - Rapid processing of scene semantics 69
  - Attention to meaning 77
  - Attentional capture and control 86

Chapter 3: Theory and General Methods 93
- Attentional Capture in Object Search by Associated Contexts 93
  - Motivating theory 93
  - Testing attentional capture 96
  - Testing capture by associated contexts 98

General Methods 100
Presentation method 100
Stimulus selection 101
Predictions for the Current Experiments 105

Chapter 4: Contextual Capture and Detection 107
Experiment 1 107
Method 107
Participants 107
Stimuli 107
Procedure 107
Design 109
Results 110
Discussion 113

Chapter 5: Contextual Capture and Discrimination 118
Experiment 2 120
Method 120
Participants 120
Stimuli 120
Procedure 120
Design 121
Results 122
Discussion 123
Experiment 3 127
Method 127
Participants 127
Stimuli 128
Procedure and design 128
Results 129
Discussion 131
Experiment 4 134
Method 134
Participants 134
Stimuli, design, and procedure 134
Results 134
Discussion 135
Experiment 5 137
Method 137
Participants 137
Stimuli, design, and procedure 137
Results 138
Discussion 139

Chapter 6: General Discussion 140
Key Findings 141

Implications for Attentional Control Processes 142
Implications for Theories of Object Recognition 149
Conclusions 153

References 154

Appendices 177
Appendix 1 178

List of Figures

FIGURE 1. Line drawings showing schematic scenes and objects from Hollingworth & Henderson (1999). 40

FIGURE 2. Digitally manipulated photographs showing a schema inconsistent object in scenes from Davenport & Potter (2004). 42

FIGURE 3. Photographic objects surrounded by associated and unassociated objects from Auckland, Cave, & Donnelly (2007). 44

FIGURE 4. The priming and target stimuli employed by Gronau, Neta, & Bar (2008). 48

FIGURE 5. A schematic temporal characterization of processing in a rapid scene classification task from Thorpe (2002). 74

FIGURE 6. Examples of the associative pairs employed in Moores, Laiti, & Chelazzi (2003). 84

FIGURE 7. Examples of the object-context photograph pairs used in the object recognition task. 104

FIGURE 8. Trial sequence in Experiment 1. 108

FIGURE 9. Hit rates for subjects in Experiment 1. 111

.

FIGURE 10. Sensitivity for object targets following associated or unassociated contexts at each lag. 112

FIGURE 11. Trial sequence in Experiment 2. 121

FIGURE 12. 2AFC accuracy as a function of the preceding contextual image and the lag condition. 123

FIGURE 13. The effects of contextual distractors and lag in Experiment 3. 130

FIGURE 14. 2AFC accuracy for object photographs in Experiment 4. 135

FIGURE 15. 2AFC accuracy for object photographs in Experiment 5. 139

Abstract

This paper describes a sequence of experiments addressing basic questions about the control of visual attention and the relationship between attention and object recognition. This work reviews compelling findings addressing attentional control on the basis of high-level perceptual properties. In five experiments observers were presented with a rapid sequence of object photographs and instructed to either detect or selectively encode a verbally cued object category. When these object categories (e.g. “baseball”) were preceded by contextual images associated with a given object category (e.g. “baseball diamond”), observers were less likely to accurately report information about the target item. This effect obtained with both detection and discrimination measures. This evidence of attentional capture is particularly strong because associated contexts typically enhance object detection or discrimination, whereas here they harmed performance. These findings demonstrate that observers use relatively abstract and elaborated representations when selecting visual objects on the basis of category. Further, even when observers attempt to ignore depictions of associated contexts these images engage perceptual processing. That is, while participants were able to determine the target of their search categorically, they had relatively little control over the specific types of representations and information employed when performing an object search task. After reviewing these five experiments, conclusions regarding the use of object-context association knowledge in vision are addressed.

Chapter 1: Visual Attention

Introduction

Understanding visual attention in natural settings requires theories that describe complex attentive behaviors (Shinoda, Hayhoe, Shrivastava, 2001). Observers seldom search sparse displays for orthographic targets, but locate and interact with objects embedded in scenes on the basis of hierarchically organized goals. In some cases the precise visual details of a target may be unfamiliar to an individual looking for an object. How is it that observers locate and selectively encode task relevant information in these perceptual tasks? What types of representations are matched against incoming sensory information? These experiments provide evidence for the hypothesis that schematized representations of contexts associated with a target item are used to guide encoding. Incoming information is prioritized to the extent it matches this rather abstract description. Most intriguingly these experiments demonstrate that once an observer has chosen a target, the way in which these guiding representations are employed is at least partly out of an observer's control. Observers can choose task relevant information, but only coarsely. That is, when observers search for a target they cannot help but attend to contexts associated with that target--even when this harms performance.

Remarkable progress has been made in understanding the basic visual properties employed by observers when selectively attending to objects or locations. Sophisticated models explain performance when observers select targets based on low-level variables such as color, orientation, or shape properties (Treisman & Gormican, 1988; Wolfe, Cave, & Franzel, 1989). On the basis of these physical stimulus properties, observer performance in a variety of attentive tasks has been described effectively. In paradigms involving visual search (Wolfe, 1998), rapid serial visual tasks (Leber & Egeth, 2006), target-distractor interference (Eriksen & Eriksen, 1974; MacLeod, 1991), and spatial cueing (Adamo, Pun, Pratt, & Ferber, 2008), performance is driven by the relationship between attentional structures, a representation of the attentive task, and stimulus properties along task relevant dimensions. These descriptions are inadequate, however, because perceptual processes typically operate on stimuli much more complex than typical visual search stimuli (Nakayama & Martini, 2010).

If one wishes to describe attentive behaviors completely, our current understanding will have to extend to treating attention to more abstract stimulus properties. In many cases observers need to identify and locate poorly characterized objects. For example, tools may need to be identified on the basis of their function or affordances (Forti & Humphreys, 2008). Targets may be identified as members of particular broad or narrow category (Evans & Treisman, 2005). In unfamiliar tasks, target identifying features may only be available after extensive training (Barenholtz & Tarr, 2007). Further complicating matters,

typically object search is embedded in a much larger task context. Few observers are looking for static objects commingled with distractors in space, but for objects and events selected on the basis of affordances from in a rich, periodic, and noisy stream of visual information. Our understanding of object search will be enhanced if research can describe the dependencies in time and space, or contextual factors, that determine search performance.

Understanding object search in context will require theories that address the sophisticated attentional control settings that inform selective processing. Attentional control settings (ACS) are representations of the criteria that identify relevant information in a selective perceptual task (LaBerge, 2002). In many tasks, for an observer to selectively enhance or exclude perceptual information there must be a guiding representation that is available for comparison to incoming perceptual information (Wolfe, Cave, Franzel, 1989). ACS may include representations of certain visual properties (Treisman & Gormican, 1988), objects (Downing, 2000), spatial locations (Posner & Cohen, 1984), search termination conditions (White & Davies, 2008), and high-level representations of a stimulus relevance (Koivisto & Revensuo, 2007). It's only due to the rich interplay of multiple ACS, and other task representation systems, that a limited capacity perceptual system can provide the information necessary to structure behavior adaptively in complex environments.

For example, suppose an observer was instructed to locate a green textbook in an unfamiliar room. Because they are familiar with some of the visual details of the book, they can preferentially attend to locations containing green

surfaces of an area roughly consistent with a prototypical textbook. In addition to this visual information, search can also proceed on the basis non-visual information about how textbooks are used and stored. An observer might know that books are often found on bookshelves or in backpacks. Observers can increase the efficiency of their search and reduce the likelihood of errors if they use all available information about the target. There is already considerable evidence that when searching for an object, observers use knowledge about the relationship between a target and locations in a scene to structure their visual exploration of a scene (Castelhano & Heaven, 2010). It is unclear to what extent the preferential allocation of attention to contexts associated with a target object is under an observer's control. The following discussion and experiments will address the predictive use of visual context in the automatic attentional selection of target objects. The selection of associated visual contexts is argued here to be automatic in the sense of being obligatory (Logan, 1992). These experiments show that when observers are searching for common objects, they involuntarily attend to scene contexts associated with those objects, even when this harms performance.

Human Visual Attention

The study of human attentive capacities has been central in the development of modern psychological theory. Generally, attention is characterized as a selective process wherein some subset of available perceptual information is sampled for more elaborated processing (Pashler, 1999). Given the range of behaviors involving the selection of information it is

not unsurprising that a number of varieties of attention have been hypothesized. For example, attentive behaviors can be considered in terms of the duration over which they operate, ranging from transient to sustained. There are many ways of further subdividing attentional mechanisms (e.g. endogenous vs. exogenous, modality-specific mechanisms, etc.). In the current discussion, visual attention will be emphasized. Specifically, visual attention will be addressed in the context of goal-directed exploratory perception with demonstrations that conceptual information is involved in the control of these attentive processes.

There is an extensive literature investigating the role of attention in the performance of visual tasks. Attention is implicated in the control of visual search (Wolfe, Cave, & Franzel, 1989), the encoding of items into visual short- and long-term memory (Potter, 1975), the memorization of spatial locations (Awh, Jonides, & Reuter-Lorenz, 1998), and a variety of other visual processes. In fact, the explanatory power of attentional theories is notable, with many models giving attentional mechanisms a central role in cognitive function. This is not surprising because visual cognition is accomplished by a limited capacity perceptual system. Demonstrations of inattentional and change blindness indicate that not all visual properties are similarly accessible (Mack & Rock,1998; Simons & Rensink, 2005). Much of the recent work on visual attention has tried to measure the relative strength and boundary conditions required for visual stimulus properties to reach awareness or guide behavior (Most, Simons, Scholl, Chabris, 2000; Koivisto & Revonsuo, 2007).

The Medium of Visual Attention

Significant theoretical development has been focused on the modes of representation used to guide visual attention. When attention selects one subset of information over another, what information is available to guide this selection? Many researchers have claimed that visual information is sampled on the basis of its location in the visual field (Graham, 1985). Alternatively, information might be selected using a perceptually organized representation of a scene (Scholl, 2001). The following discussion will briefly entertain these two possibilities. In much the same way that attention may be allocated on the using either spatial or object-centered representations, later evidence will demonstrate that attention is allocated on the basis of both visual and conceptual information.

Spatial attention models. Many early spatial selection of visual attention argued that selection occurred on the basis of location (Eriksen & Eriksen, 1974; Pashler, 1999). When attention is directed toward a location in the visual field, observers' responses to stimuli presented at that location is enhanced both in terms of accuracy and speed (Posner, 1980, Posner & Cohen, 1984). This is true both in the case of overt orienting, in which attention is directed at a location along with eye movements that fixate the location, and in the case of covert orienting, in which attention is directed toward a location in the absence of eye movements (Carrasco, Penpici-Talgar & Eckstein, 2000).

Inhibition of return is one particularly strong demonstration of visual attention operating on the basis of a spatial representation (Posner, Rafal,

Choate & Vaughan, 1985). Many researchers have presented evidence that visual attention relies on both the inhibition of irrelevant information and the enhancement of relevant information in selective processing (Watson & Humphreys, 2000). Cuing a location with an onset that precedes a target by 200 to 300 ms will facilitate responding to targets at that location. However, if the if the stimulus onset asynchrony (SOA) between cue and target exceeds 300 ms, what was once enhancement becomes inhibition. This finding is explained in terms of an observer's optimal sampling strategies while scanning visual information in the environment. It is argued that visual attention tracks and inhibits recently attended locations in order to obtain as many uncorrelated samples of visual information as possible. When the cue precedes the target by more than 300 ms, attention has already been deployed to that location and found no target. The mechanisms of visual attention then begin the search process anew at another location. This explains why responses are actually slower following a long cue to target SOA than when there is no cue at all. Researchers are able to measure the strength and distribution of this inhibition by presenting targets and onsets with various temporal and spatial parameters. The magnitude of the inhibitory costs varies directly with the distance from the cue to target, suggesting that a spatial representation is being used to track recently attended locations.

Another piece of evidence supporting a spatial conceptualization of visual selective attention involves interference paradigms in which an over-learned and automatic task is put in conflict with a controlled, deliberate task. Two examples

of interference paradigms are flanker and Stroop tasks. In flanker tasks, subjects are instructed to respond to some centrally presented stimulus and ignore immediately adjacent distractors (Eriksen & Eriksen, 1974). These distractors are associated with either a compatible or an incompatible response relative to the centrally presented stimulus (compatible- x X x; incompatible- y X y). The costs associated with the exclusion of this incompatible response can be measured in terms of accuracy or, most often, latency. The Stroop task is similar and usually involves linguistic stimuli presented in a setting where subjects must respond only to the perceptual and not the semantic characteristics of the (usually verbal) stimulus (MacLeod, 1991). When the semantic characteristics of the stimulus are incompatible with the perceptual task, or even simply engaging, costs are observed in response times. In both of these paradigms, costs associated with incompatible stimuli are reduced when the distance between the focal and interfering portions of the display is increased. Taken together, these and other data support a model of visual attention where spatial representations play a central role in the capture, deployment, and guidance of visual processing.

Perceptual organization based models. An account of selection based solely on spatial location is complicated by demonstrations that perceptual groups, or objects, can drive visual attention (Scholl, 2001). In one of the first studies to demonstrate object based attention, Duncan (1984) instructed subjects to respond to the visual properties of two spatially overlapping objects. For one object, a diagonally oriented dashed line, subjects were instructed to respond the line orientation and texture. For another object, a box with a gap along one of its

two vertical sides, subjects were instructed to respond to the size of the box and the orientation of the gap. After being shown the two objects briefly, subjects reported two of the visual properties of the objects. Importantly, these properties could be from the same object or from different objects. Accuracy was higher when the two probed visual properties were from the same object as opposed to when the visual properties were from two different objects. This two object cost was interpreted as evidence that attention can be allocated toward objects in much the same way that it can be allocated toward locations. However, it can be difficult to determine whether the two object cost observed is perceptual or mnemonic (Awh, Dhaliwal, Christensen & Matsukura, 2001). In so far as subjects are instructed to reply first regarding a property of one object and then regarding a property of a second object, two object costs could be the perceptual costs of attending to two objects in the world or the memory-based cost of retrieving information associated with two separate memory representations.

This two object cost has been replicated in experiments that control for the types of reported visual properties in a more sophisticated manner. Baylis & Driver (1993) presented observers with perceptually ambiguous stimuli containing two inward facing convex contours. Observers were instructed to indicate whether the contours matched. In a manner similar to the familiar face-vase illusion (Rubin, 1915), these two contours could be interpreted as belonging to two inward facing objects against an empty central region, or as the outer contours of a single, centrally presented object against two peripherally located empty regions. Experiment instructions biased subjects toward one of these two

interpretations. Subjects who interpreted the two contours as belonging to a single object were faster in their contour matching judgments than subjects who interpreted the contours as belonging to two separate objects. This finding is particularly interesting because the exact same stimulus and response was employed in both conditions.

Of course, few researchers would argue that all attentional selection occurs on the basis of objects. There is clearly a role for both object based and spatial mediated visual attention. The interaction between these two factors was investigated in a study by Egly, Driver, & Rafal (1994). Adapting earlier work by Posner (1980) demonstrating that observers can respond to a target faster when it is preceded by a spatial cue, Egly and colleagues presented subjects with two objects in a cuing paradigm. Targets could occur in any of 4 locations, each at either end of two rectangular objects. Distance between these four locations was equated such that the distance between locations within an object was matched to the distance between locations on two objects. Abruptly appearing spatial cues preceded target appearance in a manner demonstrated to capture transient spatial attention. The relationship between the cues and the targets was manipulated so that the cue and target could appear at the same location within an object, at different locations within an object, or on different objects entirely. As one might expect, responses to targets were fastest when the cue and the target appeared at the same location within the same object, replicating Posner and others. However, responses to targets appearing within the same object as the cue were faster than responses to targets appearing in a different object from

the cue, despite the fact that the distance from cue to target was equivalent between these two conditions. This suggests a role for both object and spatial visual attention in the detection of abruptly appearing targets within objects.

While the objects presented by Egly and colleagues were defined in terms of common region, there are other explanations compatible with the observed pattern of responses. Avrahami (1999) essentially replicated this study, but used partial object cues containing only a set of parallel lines instead of complete rectangles. A similar within object advantage was found, despite the fact that the objects were only partially indicated. Avrahami argued that the advantage for within object comparisons observed previously may in fact result from facilitated attentional guidance parallel to, as opposed to perpendicular to, the presented lines.

The possibility that the axis along which comparisons are made plays a role in object based attention was evaluated by Crundall, Cole, & Galpin (2007). Observers were presented with several dashed lines in various configurations and instructed to indicate whether two target features were contained within the same or different objects. When targets appeared along collinear portions of a given line-object group, facilitation was observed. However, when targets appeared within portions of an object that were not collinear, there was no within object advantage. The authors argue that previous studies may have conflated object based advantages with advantages due to collinearity.

Evidence gathered by Behrmann, Zemel, & Mozer (1998) extended early demonstration of object based attentional effects with objects whose spatial continuity is interrupted by an occluder. Observers were presented with objects containing either two or three bumps at opposite ends of an extended rectangle. Subjects were to indicate as quickly as possible whether the number of bumps were equivalent. The locations of the two sets of bumps were manipulated so that they either appeared on the same or different objects. A similar within object advantage was observed in both the single object continuous and single object occluded condition.

On the basis of these and other studies, one is forced to conclude that the control of visual attention occurs on the basis of both spatial properties and perceptual organization. In fact, many researchers have argued that attentional control is flexible and can be directed by a variety of modes of representation (Tipper & Weaver, 1998; Nakayama & Martini, 2010). If attention supports the guidance of action in natural environments, then object representations likely guide perceptual processing because objects are the targets of actions. However, attentional mechanisms may be sensitive to the regularities in the task and environment such that different tasks employ wholly distinct modes of attentional control.

The studies above demonstrate that visual attention is relatively flexible. In certain circumstances it appears to be guided on the basis of location in the visual field. In other circumstances, it is allocated on the basis of learned regularities of visual experience. After briefly treating some examples of models

of visual cognition where attention is given a central explanatory role in order to motivate our discussion, we'll return to discuss the nature of the attentional control settings that support this flexibility. In the following examples, attentional mechanisms are implicated many cognitive phenomena typically associated with memory (e.g. feature integration). These accounts underscore the importance of understanding the attentional control settings that govern attention in the larger context of visual cognition.

Attention in Theoretical Accounts of Visual Cognition

Feature integration theory. A broad framework for attentive perceptual processing is developed quite successfully in Treisman's Feature Integration Theory (Treisman & Gormican, 1988). This model accommodates a large body of visual cognition data with relatively simple formalizations. Stated generally, Treisman argues that performance in a variety of perceptual tasks results from the storage of independent perceptual dimensions (e.g. color, orientation, etc.) in multiple, parallel feature maps and the integration of these parallel features by visual attention.

One task where Feature Integration Theory (FIT) has been particularly successful involves the detection and identification of a target element amidst distractors, known as visual search. By presenting targets and distractors in specific combinations, researchers can measure the search efficiency of visual attention. On the basis of a large collection of search efficiency data, researchers have posited a broad distinction between the pre- and post-attentive

representations and processes that govern attentive behavior (Neisser, 1967). The attentional selection of a target in visual search is hypothesized to occur on the basis of two visual procedures: serial and parallel search. In the case of serial search, items are encoded serially into a comparison process that matches them to a top-down representation of the target. When items are searched in parallel, multiple items are compared to this target representation simultaneously. Critically, serial and parallel visual search are argued to have distinct effects on observer reaction times when plotted against the number of searched items. When observers are searching for items serially, reaction times will increase as a function of the number of checked items. When observers are checking items in parallel, the relationship between response times and the number of elements is not nearly as strong and direct as in the case of serial search. By analyzing reaction times for target-absent and target-present trials across different set sizes with varied target-distractor relationships, researchers have identified features whose processing depends on the serial allocation of visual attention and those that can be processed in parallel. Associated with these two visual routines are two modes of stimulus representation. Parallel processing is argued to operate on simple unidimensional (e.g. color) object representations whereas serial processing utilizes integrated, complex, multidimensional object files. There are a number of features such as orientation, color, and size that seem to be processed without respect to the capacity limitations typical of other visual processing tasks.

Continuing with Treisman's account, perceptual processing of featural singletons (along whatever dimension is critical in the task) occurs utilizing simple boolean feature maps that describe, in a spatially isomorphic form, the presence or absence of a given feature in a display. These maps exist in parallel and do not code object properties in relation to one another. Further, the information contained in these simple feature maps can be accessed in parallel. Apart from these individual feature maps, there exists a master map of locations. Visual attention is allocated in reference to this master map of locations. Once attention is directed to a location on the master map of locations, information at associated locations in all the separate features maps is accessed and integrated into a single object file. The process of integrating these various pieces of sensory information is known as binding. Attentive mechanisms are conceptualized as fundamentally conservative and do not invest greater resources in a given cognitive task than are required. As such, if a search task can be accomplished using one of these simple feature maps, visual attention will not be required to bind features across separate feature maps. However, when observers are looking for a target that can only be identified using information integrated between feature maps, as is the case in conjunction search, visual attention will be required to bind these separate features together into a unified percept. Once a bound representation is available, an observer can match this object file to those stored in either short- or long-term memory.

While FIT provides a wonderfully lucid description of the processes involved in visual search, it also predicts findings outside of the search literature.

Specifically, in situations where visual selective mechanisms are strained, subjects are more likely to incorrectly combine features from separate objects. These illusory conjunctions occur when information from one pre-attentive feature map is incorrectly associated within information from another feature map by visual attention. The conditions that can stress attentive performance are numerous and give a strong test to the generality of FIT account of binding and conjunction search. Since attentional selection occurs in both time and space, attention can be negatively influenced by presenting stimuli that are brief in temporal extent or broad in spatial extent. This gives attentional mechanisms a shorter amount of time per unit of visual information in which to integrate information from individual feature maps. As would be predicted, when observers view too numerous or briefly presented stimuli, their binding performance suffers. Similarly, if participants view stimuli in a dual task setting, such that less attention is available for individual tasks, illusory conjunctions are more likely. Lastly, if a participant sustains damage to the parietal cortex, an area closely associated with the allocation of spatial visual attention, illusory conjunctions occur at a pathologically high level (Robertson, Treisman, Friedman-Hill, Grabowecky, 1997). This neuropsychological disorder, known as Balint's Syndrome, results in significant disruption to object integration processes and renders victims unable to report veridical conjunctions despite viewing times lasting seconds (Rafal, 2003).

Coherence and FINST theories. An even more significant role for visual attention is described in coherence theory (Rensink, 2000). Within this account,

attention is involved not only in the binding of a durable integrated representation of an object, but is also required for the sustained existence of the object in memory. Largely drawing support from the failures of memory demonstrated in change blindness, Rensink argues that once attention is removed from an object the object, as an explicit integrated perceptual representation, ceases to exist (Rensink, O'regan, & Clark, 1997). This is thought to explain why observers experience such difficulty when detecting changes between successively presented scenes. The currently presented scene or object cannot be compared to a more durable representation in memory because a durable object file simply does not exist. Critically, only very few (4 or fewer) objects are available for attentive inspection at any given moment. Once attention is withdrawn, the object file is completely over-written by subsequent attended objects. As the author points out, this position is at variance with the modal model (Shiffrin & Schneider, 1977) wherein short term memory is an activated portion of the greater long term memory system and short term memory activation persists when attention is removed. Within the coherence model, attention is given a role beyond simple filter or binder, but is directly responsible for sustaining representation long enough for basic operations to be performed on them. It is argued that detailed visual memory is largely unnecessary because eye movements are rapid, metabolically inexpensive, and capable of delivering high quality sensory information.

The coherence model of attention fits within the larger theoretical framework of Rensink's triadic architecture. Within this model there are three

subsystems that support online scene perception. First, there is an efficient low-level visual system capable forming a volatile representation of crude visual elements rapidly and in parallel across the visual field. From this large assembly of proto-objects, limited capacity attention selects a small subset of objects for elaboration. These objects are available to awareness. Apart from the low-level visual system and the limited capacity attentional interface between low- and high-level vision, Rensink also posits a unlimited capacity high level visual attention control structure that can direct attention to various objects in the visual field according to high-level interests such as curiosity, observer motivation, or task set.

Deictic accounts of memory use indexical systems to locate and track information in the world. Rensink's indexical model shares many attributes with the fingers of instantiation, or FINST, model developed by Pylyshyn & Storm (1988). The FINST model is designed to account for performance in what the authors refer to as "situated vision"; that is, vision for the recognition of objects and the control of directed action. According to Pylyshyn & Storm, an indexical system deploys sticky pointers that track objects and their locations in the world. This tracking can occur in parallel at several locations across the visual field. It is emphasized that the maintenance of identity, or the knowledge that a certain object has a continuous and integral existence, is one of the functions of these pointers. This tracking of specific objects over time despite sensory similarity to distractors is a key component both within the FINST account and the multiple object tracking studies that provides some its strongest evidence. There are five

key elements to the FINST model as elaborated by Pylyshyn (2001). First, low-level features are segmented and clustered into perceptual similar regions, often forming objects or parts of objects. These activated clusters then compete for 4 to 5 available tracking indices. The assignment of these indices is largely stimulus driven and inaccessible to high-level considerations. This stands in contrast to Rensink's account wherein visual attention can be directed in a manner approximating top-down control. These indices, then, are bound to the available objects even if the features of those objects change over time. It is the continued identity of the object that is critical, not some set of object features. Lastly, and similar to the coherence model, only those objects which are currently indexed are available for more elaborate processing. It should be noted that initial descriptions by Pylyshyn and colleagues suggested that this indexical system is pre-attentive and tracks locations in a manner independent of attentional resources. Authors argued that some pre-attentive representation would be required to provide locations to an attentional control system that could then orient to the supplied location. However, recent evidence is consistent with the claim that the attentional tracking described by Pylyshyn draws on mechanisms shared with other visual attention tasks (Scholl, 2001).

While differences exist between Rensink's and Pylyshyn's models of attention and short-term perceptual memory performance, in both cases what is often treated as a function of memory, such as establishing object identity over time or integrating multiple visual objects, is explained in terms of a more complex attentional structure. What is primarily of interest in the current

discussion is the integral role attention is given in sustaining any sort of productive perceptual engagement with the environment. Since visual attention is something of a lynchpin for many important explanations of visual function writ large, expanding our current understanding of the processes and representations involved in attentive operations is vital. For attentive mechanisms to be ascribed such an important role in the heterogeneous tasks that comprise visual cognition, these mechanisms need to be configurable for a given task. The following discussion addresses the control of visual attention for a given task.

Control of Visual Attention

Top-down and bottom-up factors in ACS. Much effort has been dedicated to establishing the criteria available to the selective mechanisms that govern visual attention. The selection criteria, or attentional control settings, employed in a given attentive task have a variety of aspects worthy of inquiry. For example, researchers have investigated the complexity of the selection criteria. This line of research has been critical in determining whether signals can be attended on the basis of high-level attributes, such as meaning. One important theoretical distinction within these selective mechanisms involves the degree to which a given attentive act is under the control of the observer. In a variety of experimental paradigms, researchers have measured the extent to which selection mechanisms can be considered volitional or dispositional. Dispositional selection processes are usually referred to as bottom-up and involve the capture of attention regardless of the observer's intentions (although weaker formulations are permitted). Volitional selection processes are usually

referred to as top-down and involve the control of attention in a manner consistent with observer's current motivation and goals as represented in working memory (LaBerge, 2002). Recent research has focused on the distinction between these two selection mechanisms and boundary conditions for each (Theeuwes, 2004; Folk, Remington, Wright, 1994).

For example, much recent work has evaluated the role of transients in driving the enhanced processing of stimuli. Specifically, the role of onsets, or abrupt object appearances, in capturing transient attention, establishing inhibition of return, or other attentive phenomena has been a particularly fruitful domain of inquiry (Yantis & Jonides, 1996). The types of stimuli and discontinuities that can attract attention regardless of an observer's efforts remains a controversial area and is, for obvious reasons, quite interwoven with the literature establishing boundaries between top-down and bottom-up attentional effects.

At the same time, researchers have explored the nature of top-down attentional effects. It is possible to consider the nature and flexibility of attentional control settings (ACS) in a manner that is at least partly independent of general selection mechanism (e.g. guided search vs. feature integration theory). There is clear evidence that the likelihood and quality of attentive engagement with a given stimulus is at least partly dependent on the observer's intentions (Folk & Gibson, 2001). There is strong evidence that attentional control as an individual differences construct is distinct from attentional scope (Cowan, Fristoe, Elliot, Brunner, & Saults, 2006). Top-down attentional set has been conservatively defined as “a preparatory state of the information processing

system that prioritizes stimuli for selection based on simple visual features" (Leber & Egeth, 2006). It is argued that when an observer has a particular goal in mind and the behavior required to accomplish that goal has some perceptual component, temporary changes are made to the parameters of the observer's ACS. While a more specific definition is desirable, the complexity of volitional attentive behavior and the number and variety of demonstrated effects makes summary description difficult.

Before elaborating on the scope of the volitional attentive behaviors that are governed by control parameters, it's desirable to clearly identify the role of ACS within the larger attentional system. Within a given attentional module, coordinated activity is generated jointly by a controller and the controlled expression of the parameters established by that controller (LaBerge, 2002). For example, if an observer were to attend to only the green elements in a visual search display, the control module would represent the visual properties that could be used to identify the green subset of all the present elements. The controlled expression of those control parameters, if successful, would result in the selective perceptual enhancement of just those elements that fit the criteria identified by the controller. While this example involved the use of a color property, the same could be imagined for location, shape, size, or other visual dimension. Of course, this is a consideration of attentional control at the most feature-bound, detailed level. Many accounts of ACS are more expansive, encompassing higher level attentional parameters. These more abstract attentional parameters can include attentional strategies, search stopping criteria,

facilitative and inhibitory modulations of feature specific processing, or selective processing of a stimulus dimension (as opposed to a level along a stimulus dimension).

There are some perceptual tasks that can be performed with comparable levels of accuracy using distinct attentional strategies. In one such instance, observers might be instructed to detect and respond to a target with a known and unique color in a uniformly colored collection of distractors. The target can be identified either by selecting the target on the basis of its known color (feature search) or by selecting the target on the basis of its dissimilarity from the distractors (singleton search) (Leber & Egeth, 2006). It has been demonstrated that observers spontaneously develop these attentional strategies, these search strategies persist well beyond the initial experimental session, and the strategies are relatively abstract (Leber & Kawahara, 2009). In terms of abstraction, observers have been demonstrated to perseverate with a given search strategy (e.g. feature search) independent of the actual feature level required by the task. For example, it has been demonstrated that after returning to complete another RSVP task a week after initial participation, observers showed attentional capture by an irrelevant color in a manner consistent with feature based search despite the fact that the color of the target had changed between sessions.

In the case of visual search, ACS are believed to govern the termination conditions in visual search (White & Davies, 2008). When presented stimuli match observer's expectations in terms of scope (e.g. the number of to-be-processed elements), observers are less likely to report unexpected visual

elements. When certain expectations are violated unexpected elements are more likely to be identified.

Additionally, ACS can function as either excitatory or inhibitory in a given context (Watson & Humphreys, 1997). That is, these parameters can identify items for either enhanced processing or exclusion depending on task factors. If observers are presented with visual search displays that contain unpredictably colored targets but distractors of a consistent color, savings are observed because the observers are able to selectively inhibit distractors of a given expected color.

Attentional control and perceptual organization. ACS exist at many levels and include not only changes in the weights associated with different perceptual dimensions as one looks at a fixed location, but shifts of spatial attention as well. When observers shift attention either within an object or between two objects, there are a number of ways in which performance might be impacted. Brown & Denney (2007) considered the possible roles of attentional engagement and disengagement systems in between- and within-object attentional shifts. Borrowing heavily from Egly and colleagues original cuing design, the authors presented observers with a target detection task in which targets could appear either within or outside of objects. As with previous studies, the cue-target distance for both within and and outside cuing conditions was equated. By comparing facilitation when the cue lies within the same object, when the cue lies within a different object, or when the cue lies outside an object, the relative costs of disengagement and engagement can be assessed. There

was little difference when observers were required to shift attention between objects as opposed to shifting attention between an object and a location. Also, there was a larger response time cost associated with an attentional shift from an object to a location than an attentional shift from a location to an object. These findings are consistent with the idea that the within-object advantage is actually a between-object disadvantage due to the difficulty involved in disengaging attention from an initially attended object.

While traditional accounts of these object based attentional effects construe the facilitation in processing within-object features as sensory enhancement, recent evidence complicates this interpretation (Shomstein & Yantis, 2002; Shomstein & Behrmann, 2008). Within the sensory enhancement account (e.g. Desimone & Duncan, 1995), superior performance within an object is believed to result from an improvement in the quality of early sensory representations because attention has spread throughout the object. When between object competition is biased in favor of one object over another, it is able to recruit additional processing resources for all contained features. An alternative account, suggests that object structure influences the prioritization of subsequent attentional samples. When an observer is instructed to make judgments regarding multiple features within an object, it is easier for attention to sample information within, as opposed to between, objects. Shomstein & Yantis (2002) presented subjects with a flanker task in which the distractors were either contained within the same object or contained in a different object. Regardless of the grouping of flanking distractors, interference effects were similar. Only

when observers were presented with a task that required the exploration of a given object (due to spatial uncertainty regarding target position) did an object based advantage obtain. The authors highlight the difference between the valid cue-same object and invalid cue-same object conditions in the study by Egly and colleagues. They argue that if attention spreads completely through the object, there would be no difference between these two conditions. However, the opposite pattern was observed with a valid cue advantage, suggesting that attention does not spread equally through an object.

Shomstein & Behrmann (2008) explored two possible means by which attentional prioritization might occur. On the one hand, prioritization could occur on the basis of configuration such that objects defined in terms of gestalt grouping principles (e.g. common region) are preferentially explored and sampled. In contrast, prioritization could occur using probabilistic guidance such that targets at high probability locations will be detected before targets appear at low probability locations. The rationale for this second possibility is supported by the fact that cue validity in the study by Egly and colleagues was manipulated so that following a cue, targets appeared within the same object on 87.5% of trials. The existence and relative strength of these two possible prioritization mechanisms was explored in a two rectangle cuing paradigm similar to Egly et al. (1994). Subjects were instructed to report the identity of a target letter (either a "T" or an "L") located at the end of one of the two rectangles. The strength of the configural grouping was manipulated by previewing the two objects for either 200 or 1000 ms prior to cue. The target-cue probabilities were manipulated so that

invalid different object cues were highly probable and invalid within object cues were less probable. Longer preview times resulted in larger object based effects. These object based effects can be attenuated in situations in which cue-target probabilities are such that within-object prioritization affords no advantage for observers. Taken together, the authors argue that previously demonstrated within-object advantages are likely due to both probabilistic guidance and configural cues.

The advantages that accrue for comparisons within objects have an analogue that appears when observers are instructed to make comparisons within an object part, as opposed to across objects parts. Barenholtz & Feldman (2003) presented observers with patterned objects consisting of repeated curve segments. Each object could be divided at concave curvature minima into several equivalently sized parts or regions. Subjects indicated whether two small marks that appeared along the contour of these objects contained the same or a different number of peaks. Subjects responded more rapidly when the two marks were presented along a contour of an object part. When the two marks appeared on different parts of a given object, responses were not as rapid. It should be noted that the contours were chosen such that within and between part comparisons both involved a comparison across an equivalent curve. These results inform and complicate our previous discussion of object based attention. Specifically, they show a within-object within-part advantage, extending previous findings to groupings within objects.

Taken together, there are four main explanations of within-object advantages (Brown & Denney, 2007). Within the biased competition account (Vecera & Farah, 1994; Desimone & Duncan, 1995), when objects compete for representation, selection of one object in given a perceptual system, biases other systems to represent that same object. When an observer attends to an object and responds to multiple features within that object, there is no need to change the top-down biasing signal that affords advantages to that object. When an observer is forced to respond to features contained within two objects, a shift of attention between objects requires a new biasing signal to identify to the second object. This results in a two object cost. On the other hand, the prioritization account proposed by Shomstein and colleagues suggests that the within object advantage is the result of an inherent bias within the attentional system such that when a stimulus must be explored for features (due to positional uncertainty), the system checks locations within the currently attended object before shifting to a new object. The attentional guidance account proposed by Avrahami and others argues that attention spreads automatically within perceptually organized structures, regardless of their object status. In the final account presented by Brown & Denney (2007), object based effects are not due to a within-object advantage, but rather are the result of a between-object disadvantage. When a comparison spans objects, attention must disengage from one object before engaging with a second. This disengagement process is effortful and harms performance.

There are situations in which the grouping of stimuli harms perceptual performance. Rensink & Enns (1998) investigated the nature of preattentive grouping in visual search tasks. Observers were presented with a shape combination search task in several conditions. Targets either contained two shapes ordered in depth (with partial occlusion) or with a narrow strip of empty space between the two shapes. When observers searched for the depth-ordered shapes, search was effortful and serial. When the shapes were slightly separated, detection was facilitated, arguable due the increased distinctiveness of the shapes (the contour of the shapes in the absence of an occluder was more complex). On the basis of these data, the authors argue that shape representations overcome occlusion in a rapid preattentive manner, and the subsequent search involves an exploration of the display space for a shape that is similar to distractors (due to the completion process). Once these shapes are "filled in" by this completion process, access to their constituent features in limited.

Similarly, Davis and colleagues present evidence that the within-object advantage observed in the literature can be reversed in situations wherein the single large object contains more perceptual information than two smaller objects (Davis, Driver, Pavani, & Shepherd, 2000; Davis & Holmes, 2005). Additionally, a modulation of object based effects was observed depending on the onset relationship between the object and the to-be-discriminated features. When the features appeared at the same time as the objects, observers were actually faster comparing features between separate objects. On the basis of these and

other related findings, Davis modifies the argument presented by Humphreys and Heinke (1998) and argues that: a) individual features are bound within objects using links based on parvocellular processes whereas features between objects are related on the basis of magnocellular processes and b) the number and strength of these within and between objects links, rather than the number of attended objects, underlies the within and between object effects observed. Manipulations of the distribution of information across spatial frequencies with concomitant changes in object based effects support this interpretation.

Several recent studies by Moore and colleagues have explored the role of perceptual organization in the control of the visual selection of objects. These experiments draw on the attentional walk paradigm developed by Intriligator & Cavanagh (2001). In this paradigm, subjects are presented with a collection of circular elements (disks) organized into a ring. At the start of a trial, one of these element is indicated to be a starting point. Following this initial cue, participants are given instructions to move the focus of their attention one element over to either the right or left. This continues for some period of time. Following this sequence of directional instructions, observers are tested to see how accurately they could individuate and select the target disk indicated by the initial cue and subsequent shift cues. The spatial resolution of visual attention is assayed by manipulating the density of the disks. Initial interpretations of performance limitations identified attentional resolution as the limiting factor in tracking performance. Recent evidence indicates that performance is more likely limited by the precision of attentional control (Moore, Hein, Grosjean, and Rinkenauer,

2009). The role of perceptual organization in the implementation of attentional control was investigated by Moore et al. (2009) in the following manner. Subjects were presented with disks arrayed in a ring in alternating colors. Depending on the presented tone, subjects were instructed to shift attention to either the left or right disk of the same color. Because the disks were in alternating color order, this would involve a shift of two disks. If the identically colored disks were grouped, and attentional selection occurs within these groups, then performance with 24 disk in alternating color order should be comparable to performance with 12 disks. Despite multiple grouping manipulations, including connecting like colored elements and organizing binocular displays so that like colored elements fall along the same depth plane (different from the depth plane of the dissimilarly colored disks), researchers found little evidence that perceptual organization could be used to guide attentional selection of the target disk. The authors argue that the same grouping principles that facilitate attentional performance when observers are instructed to respond to multiple stimulus attributes can hurt performance when observers are required to individuate items.

Regardless of whether the effects of perceptual organization harmed or helped a particular attentive task, one can conclude that attentional control mechanisms are governed by systems that incorporate knowledge about the perceptual organization of one's visual experiences. Recent accounts of perceptual organization in scenes have emphasized the role of repeated exposures to patterned stimuli in the development of grouping rules (Fizer & Aslin, 2001; Fizer & Aslin, 2005). The novel experiments presented in this paper

will investigate the role of abstract associative knowledge in the control of visual attention and demonstrate, in a manner analogous to these demonstrations of the influence of perceptual organization, that conceptual organization informs attention.

Locus of attentional selection. Within the attentional literature a number of theoretical distinctions have been drawn identifying common attributes and distinct features of hypothetical attentional mechanisms. Theories of attention can be organized a number of different ways, including the evaluated sensory modality, the selective mechanism and the locus of selection. The following discussion will focus on the last of these distinctions.

Much early research focused on where in an information processing sequence attentional selection occurred. Because information processing models of human cognitive performance emphasized the sequential operations involved in generating mental behavior and the capacity to conduct these operations is limited, attention was hypothesized to operate within or between distinct stages of cognition. Indeed, in so far as distinct cognitive stages have associated modes of representation, much discussion centered around the types of perceptual representations and attributes that are available for attentional selection. If the role of attention is to select the most important information available for elaborated processing following some sort of perceptual bottleneck, where in processing does this bottleneck emerge?

Broadbent's Filter model argued that perceptual and cognitive processing of a stimulus involves three distinct mechanisms (Broadbent, 1958). First, information is registered in a high capacity, volatile sensory representation often referred to as iconic or sensory memory. Information here is rich, but close to sensation and not elaborated in terms of meaning, goals, or other high-level factors. Next, information is shunted from this high capacity buffer through a filter which permits only a portion of the available information through. Attentional selection is fundamentally dichotomous, with selected information passed along unaltered and non-selected information ignored entirely. Unattended information is unavailable for later, more elaborated treatment. Further, he argued that the criteria available to these selective processes operated on largely early, sensory attributes. On the basis of these physical properties alone, some information is allowed to pass and other information left unprocessed. If true, this meant information could be selected on the basis of pitch, brightness, or loudness, but not its meaning. Sensibly, Broadbent argued that, in order for attention to effectively pare down the mind's computational burden, selection must occur before these limitations are manifest. Performance limitations were thought to become greater later in processing, so selection likely occurred earlier in processing. Following filtration, information is passed along to a detector which processes the information in a manner consistent with the stimulus' meaning, the observer's goals, etc. Because selection occurs early in processing, this class of models, as typified by Broadbent's filter model, are referred to as early selection models.

Early selection models have the advantage of making strong predictions regarding the fate of unattended stimuli . Specifically, information about properties of an unattended stimulus that are thought to not emerge until late in processing, such as meaning, should be unavailable to the perceiver. This prediction was disconfirmed by Moray (1959) is his famous demonstration of the "cocktail party effect." Observers were presented with a auditory stream containing a narrative they were to shadow while various types of information were presented in the unattended channel. In one condition, observers were presented with a list of words that repeated 35 times. Despite the numerous repetitions, they were unable to recall any of the words presented in the unattended channel. Contrastingly, when observers were presented with their name in the unattended channel, they were able to recall this at the end of a given trial. This suggests that, at least in certain circumstances, observers are able to process the meaning of an unattended message.

In order to accommodate this and related results, early selection models were modified to permit certain types of more advanced operations to be performed on unattended information. At the risk of circularity, it should be clear that if the role of attention is to select some information for more elaborate processing while excluding other information, not all information is processed to the same extent. Just how elaborate the processing of unattended information is has been studies extensively. Models that permit advanced processing of unattended information are referred to as either intermediate or late selection models, largely depending on the complexity of unattended processing. Models

that suggest only partial or incomplete processing of unattended stimuli, such at Treisman's Attenuation model (1964), are referred to as intermediate selection models. Descriptions of attentional processing which assert that unattended information is processed in a fairly elaborate manner are identified as late selection models (Deutsch & Deutsch, 1963).

Considerable evidence has accrued that the meaning of unattended stimuli is accessed to some extent (Treisman, 1960; Moray, 1970; Lewis, 1970; MacKay, 1973). The degree to which these stimuli are processed, however, remains unclear. Kahneman & Treisman (1984) argued that discrepancies within the literature regarding the locus of selection reflect fundamentally different conceptualizations about what attention does. Specifically, older research was grounded in a filtering paradigm which emphasized the exclusionary role of attention. Demonstrating these effects often involved difficult tasks involving a number of cognitive systems which strained attentional control capabilities. In contrast, recent work takes a selective set paradigm and generally involves simpler tasks. This may make consolidating the contradictory literature on the fate of unattended stimuli difficult.

Recent evidence suggests that the degree to which unattended stimuli are processed depends on perceptual load, or the amount of information attention must filter (Lavie, 1995). According to this perspective, attentional selection is only necessary when the amount of presented information exceeds the capacity of a limited bandwidth channel. Studies which find processing of unattended information to a greater or lesser extent do so because the amount of information

presented exceeds the capacity of this channel by some varying amount. Interference paradigms, such as the Stroop task, are one way of measuring the extent to which unattended stimuli are processed. Numerous studies have demonstrated that usually automatically processed, but unattended, stimuli are less likely to exhibit an interfering effect when perceptual load is high (Lavie, 1995).

Conclusions

This chapter provides a context for the following discussion of attentional capture by scenes related to a target object. We have sampled perspectives that emphasize the role of attentional resources in the consolidation and elaboration of detailed object and scene representations. These attentional resources can be directed in sophisticated ways that reflect an observer's experiences. The conceptual organization of objects and scene knowledge is hypothesized to direct visual attention in much the same way that the spatial organization of objects and scenes. Specifically, the regularities of this knowledge is argued to direct and limit the allocation of visual attention during object recognition. The following chapter will elaborate this hypothesis, showing its continuity with theories of contextual influences on object recognition.

Chapter 2: Object-Context Associations in Object Recognition

The current experiments test the hypothesis that observers involuntarily attend to contexts associated with the current target in an attentive task. Of course, this claim has consequences for both our understanding of attention and object recognition. This chapter will focus on the relationship between object recognition, object selection, and visual context.

Direct Measures of Contextual Influences on Object Recognition

There is considerable evidence that contextual associations play a key role in object perception (Palmer, 1975; Biederman, Mezzanotte, & Rabinowitz, 1982; Bar & Ullman, 1994; Hollingworth & Henderson, 1999; Davenport & Potter, 2004). The role of visual context in object recognition has been a contentious issue. Object recognition in general is poorly understood, so it is not surprising interacting factors are disputed. The following discussion contains a sample of recent evidence and theory regarding contextual influences in object recognition, focusing primarily on the role of schematized context in object recognition. The visual information that is typically encountered along with a focal object, including associated objects and typical spatial configurations, has been shown to influence recognition performance in a wide variety of object recognition tasks. This associated information has been referred to by a wide variety of terms including context frame, schema, gist, or scripts. The same terms are employed

similarly to explain the role of pragmatic assumptions in psycholinguistics (Henderson & Ferreira, 2004). In both cases the interpretation of an underdetermined stimulus, either an utterance or a scene, is constrained by a perceiver's prior knowledge. Because of the range of terms and rather loose definitions, in this paper I'll refer to associated schematized contextual depictions as either an associated context or a related context.

For example, a schematic scene category might be "kitchen". A kitchen has typical visual properties, associated objects, and associated activities. One would expect to see an oven but not a fire hydrant. How these expectations shape selective processing of target items when the target's identity is known in advance is uncertain. Because human object recognition is performance is so effective, researchers typically have to either retroactively probe a degraded visual stimulus or measure the response time required for an observer to make some judgment about a stimulus. The following examples fall under these two broad methodological categories. However, generating models solely, or even primarily, on the basis of these two types of data will limit our understanding of the role of visual context for reasons described later.

In their influential papers, Hollingworth & Henderson (1998,1999) identify three possible relationships between contextual and object information during object recognition. In each of the three cases, it is assumed that the scene is initially characterized in some sort of crude feed-forward initial categorization before the processing of individual objects proceeds. There is considerable evidence that scene category information is available early in visual processing

(Greene & Oliva, 2008), possibly on the basis of low spatial frequency information (Bar, 2004). In the first possibility, the contextual information is able to enhance the description of the target item, resulting in a genuine increase in sensitivity. In this case, the features of the object that are diagnostic in the observer's task are encoded more rapidly, less susceptible to interference, or easier to recall in the presence of associated contextual information. Alternatively, observers might simply modulate their standard of evidence depending on the context surrounding the item. This biasing effect is a reasonable strategy in a world containing correlated objects and features. In this case, observers would be more likely to recognize the refrigerator only because the associated context reduces the observers' standard of evidence. Observers still recognize objects in associated contexts more quickly and accurately, but would do so only because of their positive bias. Lastly, it is possible that a scene has no influence on the recognition of objects. In this functional isolation account, object recognition and scene recognition are largely independent processes.

Hollingworth & Henderson (2000) evaluated these possibilities by presenting observers with line drawings of common objects placed in drawn scenes. The scenes were presented briefly and observers indicated whether particular items had been present in the initial display in a two alternative forced choice (2AFC). The authors were careful to include trials where either both or neither of the two options provided were consistent with the overall scene. If observers generate better mnemonic descriptions of objects appearing in an

associated context, they should be able to identify which of two schema-consistent items was present in the display. Alternatively, if observers are simply guessing on the basis of the overall scene schema, there should be false alarms for consistent items that are paired with inconsistent items in the forced choice following viewing. Observers showed no advantage for semantically consistent objects when discriminating between pairs of consistent objects. In fact, observers were actually more accurate when they were recognizing schematically inconsistent objects. This is likely due to the fact that unexpected objects attract attention and receive more encoding. The authors conclude that object recognition is isolated schematic scene constraints and previous demonstrations of such an effect were likely due to strategic guessing.

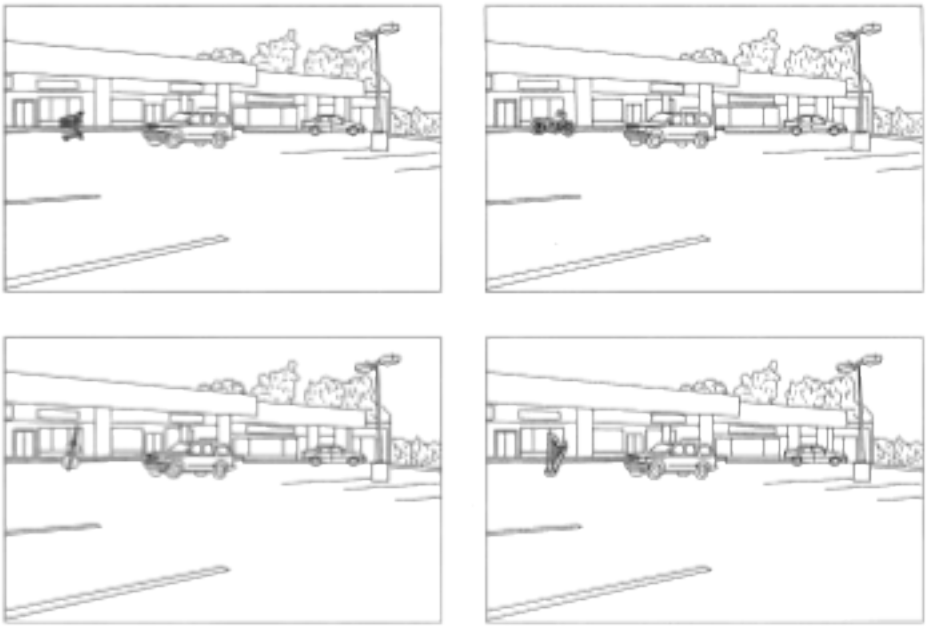

Figure 1. Line drawings showing schematic scenes and objects from Hollingworth & Henderson (1999) (Figure 1, p.325)

Davenport and Potter (2004) provide evidence consistent with description enhancement through object-context consistency in an object naming task. Observers were presented with digitally manipulated photographs containing either a single central object against a consistent or inconsistent background. Immediately following a brief presentation of the photograph, observers were instructed to indicate the foreground object(s) or background with an open-ended response. Observers identified the target item more accurately when it was consistent with the background. For example, it was easier for observers to recognize a priest against a church interior background compared with a football field. In a later experiment, Davenport (2007) demonstrated that this advantage for schema consistent objects did not depend on the number of foreground objects (1 or 2). Additionally, in much the same way that context influenced the identification of objects, the foreground objects influenced the identification of the background. Davenport advances an interactive account, similar to Bar's (2004), where centrally presented objects and the scene background act as mutual constraints during dynamic cue extraction process. This interactive model is in many ways analogous to the interactive-activation model of word recognition (McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982). In this model of letter and word recognition, word, letter, and feature information are each processed simultaneously as partial information at each level is used to constrain and inform the search at other levels. In much the same way, the foreground objects and background in the photographic scenes together determine an

integrated percept that is available for subsequent report. Consonant with this account, there is evidence that inconsistent objects can affect scene categorization quite early in processing (Joubert, Fize, Rousselet, & Fabre-Thorpe, 2008).

Figure 2. Digitally manipulated photographs showing a schema inconsistent object in scenes from Davenport & Potter (2004) (Figure 1b, p.561)

Auckland, Cave, & Donnelly (2007) provide evidence of criterion modulation in an object recognition task embedded within flanking photographic distractors. Observers were presented with an array of visual objects on each trial and then instructed to identify a target object in a 6 alternative forced choice discrimination. Target objects were photographs of common items presented very briefly in the center of a square formed by the presentation of four additional flanking objects. Together the central object and flankers formed a quincunx. If the centrally presented item was a photograph of poker chips these flanking objects could be schematically consistent (playing cards) or schematically inconsistent (grapes) with the target item. These additional photographic objects preceded the target object by either 104, 52, 0 ms, functioning as a variable onset prime of a sort. After this lead time, the central object and four flanking

objects were visible for 52 ms, followed immediately by a six alternative forced choice discrimination. The lures, or non-target items, in the forced choice on each trial were chosen to share either perceptual or semantic features with targets. Additional choices were presented to prevent observers from strategically guessing the target on the basis of the options listed. Observers were more accurate in their identification of the centrally presented object when it was preceded by surrounding schema consistent items. This shows that the presentation of associated items enhances the perception of a target item, as has been shown in variety of other context cueing experiments. However, only when the context preceded the target item did these benefits obtain, suggesting that the contextual information is most useful when the observer is getting ready to encode the target item. In this sense, the results support a criterion modulation account. It's not the case that observers were actually able to see schematically consistent items better in the presence of related information or performance for the objects sharing a common onset with the distractors would have been influenced. When observers were presented with contextual items before the target, observers did better with consistent items.

Figure 3. Photographic objects surrounded by associated and unassociated objects from Auckland, Cave, & Donnelly (2007) (Figure 1, p. 333)

Indirect measures of contextual influence. Each of the previous three examples measured object recognition performance in the presence or absence of associated contextual information. The relationship between objects and associated contexts can be tested in other, less direct, ways. The following examples describe the influence of object and context relationships on the allocation of visual attention without an overt object recognition task.

In recent demonstrations of the importance of object-scene associations in the deployment of visual attention, Gordon (2004, 2006) presented subjects with line drawings of natural scenes for a range of very short durations. Located within these scenes were objects that were either consistent or inconsistent with the schema of the scene. Immediately after the scene offset a single spatial probe appeared, and subjects were to respond as quickly as possible. If the schematic relationship between an object and the contextual scene influences visual attention, one would anticipate differences in response time depending on whether the spatial probe appeared behind a consistent or inconsistent item (but see Schmukle, 2005 for discussion of dot probe reliability). For inconsistent

stimuli, only one object would violate schematic expectations. In order to ensure that subjects did not strategically attend schematically inconsistent objects on trials containing schema violations, catch trials were included. Catch trials, where the spatial probe appeared at fixation, represented 1/3 of all trials. This ensured that observers did not strategically attend to either consistent or inconsistent items. On all non-catch trials, the probe was placed at the location of a schema consistent or schema inconsistent object. Gordon measured the time required for observers to respond to the spatial probe. He assumed that observers would respond to the spatial probe more rapidly when it appeared behind the current focus of their attention. Further, by manipulating the exposure duration of the scene, differences in attentional allocation can be measured as they evolve in time.

An inconsistent object advantage first emerged approximately 150 ms after stimulus onset. That is, when the scenes contained an schema inconsistent object, it wasn't until the scene had been visible for 150 ms that observers showed different response times towards spatial probes appearing behind schema consistent or schema inconsistent items. Because these scenes were presented for durations under the 200 ms required to plan and execute a saccade, inconsistent objects must be identified and treated differently from consistent objects in a single fixation. This does not necessarily imply that inconsistent objects are processed with any priority. Rather, Gordon suggests that after schema consistent objects are identified, resources may be allocated to schema inconsistent objects.

The allocation of attention to consistent and inconsistent objects was explored further in a negative priming paradigm. In this experiment, subjects were presented with scenes with consistent and inconsistent objects immediately followed by a stem-completion task. In stem-completion tasks, observers are presented with an incomplete word (e.g. “toa_ _ _ _”) and instructed to fill in the blanks to form the first word that comes to mind that is consistent with the completed portion of the word. In this experiment, subjects were less likely complete a word stem with a consistent object if the scene with the consistent object also contained an inconsistent object. For example, if observers were presented with a kitchen scene containing a fire hydrant, they would be less likely to complete the item “toa_ _ _ _” than if observers viewed a kitchen scene with no inconsistent items.

This prioritization of semantically inconsistent objects within one fixation is replicates previous research indicating that semantically inconsistent objects are generally more likely to be attended. Hollingworth & Henderson (2000) found that subjects were more likely to detect changes in a change blindness paradigm when the changes were made to semantically informative objects. Objects were considered semantically informative when they were inconsistent with the schema for the scene. Semantically consistent objects are not considered informative individually because they all have equal diagnostic content. Further research demonstrated an inconsistent object advantage even when the influence of eye movements were controlled, suggesting that inconsistent objects are preferentially encoded. Within the memory schema hypothesis proposed,

semantically consistent objects are set to their default schema value, but inconsistent objects are stored more accurately. Regardless of the specifics of scene semantic relation processing, this finding is helpful in our discussion of contextual influences in object recognition. These data indicate that in as little as 150 ms the semantic properties of many illustrated scene objects can be extracted, a scene can be categorized, and the locations of schema-inconsistent objects can be selected.

Gronau, Neta, & Bar (2008) demonstrate both the importance of object-context relationships and provide some hints about possible neural mechanisms in an object recognition priming study. Observers were instructed to indicate on each trial whether an image represented a real or imaginary object. This task is considered an indirect measure of object recognition because observers were not instructed to make a decision regarding the actual identity of the target object, but simply to determine whether or not it existed. Immediately preceding the presentation of this target item, a priming object was presented. The associative relationship between the priming object and the target object was structured so that the prime was either related or unrelated. Further, the specific configuration in which these items were presented was designed to be either random or reflect the typical configuration of the items. For example, if the target object and prime were related and presented in a typical configuration, one might see an oven followed by a pot immediately above it. Spatial and semantic congruency were manipulated factorially. Subject responded faster to true objects when they followed semantically associated prime objects at a spatially

typical location. The interaction of semantic and spatial congruency conditions is argued to support an account wherein contextual influences in perception arise through context frames (Bar, 2004). Described from a cognitive perspective, context frames are hypothesized to be integrated representations of objects and associated contextual details that are derived from experience. These prototypical representations of particular contexts (e.g. kitchen) include information about a scene's associated objects, typical configuration, and regular activities. Hemodynamic activity monitored via fMRI indicated that activity was concentrated in the inferior prefrontal cortex and lateral occipital cortex while subjects completed this object discrimination task. These two areas figure centrally within Bar's larger model of the influence of context on object recognition

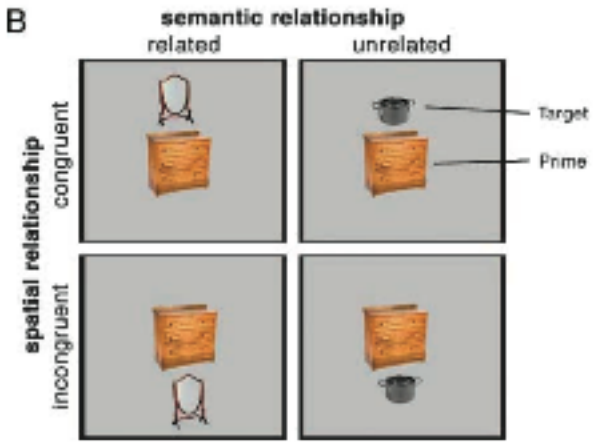

Figure 4. The priming and target stimuli employed by Gronau, Neta, & Bar (2008), (Fig 1b, p.375)

A model of object recognition using contextual information. Bar (2004) has developed a detailed neurocomputational model to describe the influence of context on object recognition with extensive psychophysiological and behavioral support (Bar & Aminoff, 2003; Bar, Aminoff, & Schacter, 2008). Within this interactive account of object recognition, when observers glance at a scene, an initial lowpass characterization of the image is propagated rapidly and in parallel across the visual field through fast acting magnocellular visual pathways. This lowpass scene is then categorized in the inferior prefrontal cortex (iPFC). This process is efficient and seems to utilize global, statistical properties of the scene (Greene & Oliva, 2009). This initial guess of the scene category is coarse, but can begin to bias object recognition processes towards outcomes consistent with the scene category. At the same time as this scene category processing is going on, observers are using ventral visual pathways to resolve object details. These scene and object categorization processes are going on I qn parallel under conditions of mutual constraint. The neural interface that mediates this interaction is located in the parahippocampal cortex. These areas are long associated with episodic, layout, and scene memory. Two specific parahippocampal regions, the parahippocampal place area and retrosplenial cortex, are argued to mediate the interaction between the iPFC and candidate object representations in the inferior temporal cortex.

Bar's model efficiently summarizes a wealth of object recognition data, but does not make clear predictions in situations where task information specifies relevant objects or regions in a scene in any detailed sense. Itti & Arbib (2005)

describe a conceptual model of object and scene comprehension with a significant role for these task factors. Within the Salience, Vision, and Symbolic Schemas (SVSS) account, interactions between long-term memory (LTM), short-term memory (STM), and various topographic representations of a two dimensional scene support complex, linguistically focused scene cognition. A more detailed description of the model is presented following the next section.

In each of the previous experiments, the influence of context on object recognition was assessed using an object recognition task where the target category was not known to the participants in advance. In the present experiments, observers will be searching for a category specified on a each trial. The way in which contextual associations might influence object recognition in these experiments is less certain.

Contextual Associations in Visual Search

In the previous section's studies, observers' ability to rapidly identify an object was influenced by scene context. Observers were required to either identify an object out of a list of alternatives, to freely recall the name of an object, or identify an image as an object or non-object. In none of these cases are observers looking for a precued object. Since the present experiments addresses the role of visual context in visual search for a known target, rather than recollection, let us turn our focus to object-context associations in visual search.

Contextual cueing. Demonstrations of contextual cueing measure the gradual build up of contextual associations regarding the location of a target object. Chun & Jiang (1998) presented observers with a visual search task containing arrays of oriented letters. Unbeknownst to the observers, some of the visual search displays were presented multiple times. While performance improved for both novel and repeated displays, there were savings beyond general task learning associated with particular repeated displays. Contextual cueing is generally assumed to reflect the gradual extraction of local configural properties surrounding the target (Brady & Chun, 2007) although recent accounts pointed towards decisional factors (Kunar, Flusberg, Horowitz, & Wolfe, 2007; Schankin & Schubo, 2010). A configural representation is argued to guide attention towards locations where targets appeared previously. Intriguingly, observers are often unaware of the repetition manipulation and cannot distinguish novel from repeated displays in an explicit recognition test.

While generally approached using sparse displays of basic elements, recent demonstrations of contextual cueing have used real-world photographic scenes (Brockmole & Henderson, 2006). Observers were presented with photographic scenes containing an small oriented letter. Some scenes were repeated while others were only presented once. Observers demonstrated faster responses to targets located in repeated scenes compared with novel. In this case, observers were able to explicitly recognize the repeated scene. Observers demonstrated a smaller contextual cueing effect when the scenes were inverted.

Apparently the more meaningful background of the upright stimuli allowed observers to retain target locations more successfully.

This semantic effect in contextual cueing was explored further by Goujon, Didierjean, Marmeche (2007). Observers were presented with a numerical search task in arrays containing a distributed set of arabic numerals. In typical contextual cueing experiments the overall or local configuration of items predicts the target location. In these experiments, properties of the numbers which comprised the display predicted the target location. It was demonstrated that the repetition of particular numbers or number category ("odd" vs. "even") could be associated by the observer with a particular target location. It is important to emphasize that the parity of the presented numbers, not their spatial configuration, that predicted the target location. However, as with traditional contextual cueing, observers were unable to verbalize this target associated information.

Models of contextual influences on scene search. Torralba, Oliva, Castelhano, & Henderson (2006) investigated how these contextual associations might be used in natural scene exploration. Within Torralba and colleagues contextual guidance model, observers' eye movements are guided by both local saliency and priors based on previous exposures to the target object. For example, if observers were searching for a pedestrian in an image, it would be reasonable for them to be located somewhere along the ground plane. The location of the ground plane is extracted through a holistic process that detects statistical regularities in the image and generates candidate locations. This map

of candidate locations is then integrated with an activation map based on local salience. This combined representation is used to direct eye movements within a scene. When observers are instructed to visually explore photographic scenes for verbally labeled objects, their fixations are predicted by the spatially licensed, schematically expected locations of the target objects (Torralba, Oliva, Castelhano, & Henderson, 2006; Malcolm & Henderson, 2010).

Taken together, these studies indicate that observers are able to deploy visual attention in a way that reflects the regularities of their experience. Participants' knowledge about probable target locations enhanced performance in these search tasks. However, scene knowledge is not limited to the patterned location of targets in visual search displays, but includes information about more complex scene relationships. One account of how general scene knowledge might be employed to answer particular questions about a scene is provided by Itti and Arbib (2005). As mentioned previously the SVSS conceptual model describes how task knowledge might inform high-level scene perception. Schematic knowledge in LTM is used to generate and evaluate hypotheses about a scene that are used in generating verbal scene descriptions. While the model describes both how observers might answer questions about a visual scene and how they might describe it, we'll only focus on the way that incoming sensory information is evaluated for relevance against a task representation.

Central to the Itti and Arbib account of scene cognition is a construct known as the minimal subscene. The minimal subscene is comprised of those elements within a scene judged to be relevant to the current scene task. This

subset of objects are evaluated relative to a central anchoring element in the scene. This anchoring element can be an object, agent, and action. The particular minimal subscene constructed by a given observer will be determined by his or her task. Not only would different observers construct different minimal subscenes while viewing the same scene with different goals, the minimal subscene of a given observer changes over time as additional scene information is extracted and goals are refined. This minimal subscene exists in short-term memory and acts as an interface between long-term scene knowledge and various spatial representations of the scene. The minimal subscene is hypothesized to play a central role mediating concrete perceptual representations and symbolic/linguistic representation of the scene.

The control of this minimal subscene extraction process is grounded in models of distributed, schematic control processes. A library of schemas are stored in long-term memory and sampled on the basis of task representations. Then, these active schemas interact within short-term memory to generate descriptions of the scene. A variety of schematic description mechanisms operate simultaneously and cooperatively as observers sample objects and features for the minimal subscene. These schemas are complex and contain not only mechanisms that represent visual features, but also control the feature extraction process, assert claims regarding other regions of the visual field, and maintain a confidence level regarding assigned labels.

Given the range of possible scenes and the vast amount of schematic scene knowledge in LTM, optimal scheduling of these schemas becomes a

thorny computational issue. Here Itti & Arbib suggest attention has a central role in scene understanding. Attention not only controls the prioritization of locations in a visual task, but also the types of features and objects to be attended in those locations. In this sense, attention directs not only the sampling of schemas from LTM but also the organization of the sampled schema instances in STM. However, the control of this cooperatively computational process is distributed and emerges only through the interaction multiple simultaneously active object schemas.

The iterative steps of the SVSS model are as follows. When observers plan on answering a particular question about a scene, there is a preparatory task biasing which prioritizes certain types of visual information before the scene is viewed. Once the scene is visible, observers extract features from the scene and construct on the basis of these features a verbal label for the scene (gist), salience maps describing local feature contrast, and task-relevance maps that prioritize scene regions on the basis of observer goals. This feature extraction and activation map construction process, is followed by recognition of particular items within the display. Once an item has been recognized, observers update scene representations to incorporate this additional information.

Preparatory task biasing primes or sensitizes both perceptual and conceptual representations in STM. This sampling of schemas from LTM into STM is accomplished on the basis of both explicit task instructions and associations in long term memory. For example, if an observer was instructed to determine whether a scene contained a goldfish, they would sample not only

goldfish schemas from LTM but also fish tank schemas. Once these candidate objects are active in STM, they can be compared with particular items and regions in the sequence determined by the overall attention activation map.

The overall activation map in turn is determined by the integration of a task relevance map (TRM) and salience maps. The TRM describes the regions of the scene that are germane to observer's goals as they view the scene (Navalpakkam & Itti, 2003). The TRM integrates information regarding the target's local features, the scene gist, and the scene layout into a single topographic activation map. For example, if an observer is looking for humans walking on a beach, activation will be higher for regions containing roughly elliptical shapes with a vertical primary axis. Knowledge of the category of scene (beach) and the layout of the scene (looking out into the water) would direct resources towards locations in the lower visual field near the shoreline. The TRM that integrates these pieces of information is then combined with a salience map to form a final map that guides attention. The TRM does not contain detailed information about the objects found in these locations, but simply a prioritization of different regions that can be used to schedule detailed visual analysis.

The SVSS model describes how observers might extract task relevant information from complex visual scenes using a collection of schematized object descriptions under distributed control. This distributed control system manages the scheduling of perceptual tasks using both perceptual and conceptual knowledge. For the current experiments, this model would suggest that when

observers are searching for a known target item, they use a complex of schemas which likely includes information about associated contexts. The model does not make detailed predictions regarding the extent to which this schema selection process is under the control of the observer. Can observers select a narrow set of schemas that prioritize only information closely associated with the target and task or is this process automatic? Itti & Arbib suggest that the schema for a target not only includes where to look but also how to look. To what extent is this latter type of schematic information sensitive to a particular task? That is, if observers are looking for an object and are presented with an associated context that, in the experimental setting, never contains the target object do observers still attend to this associated, schematic context? The current experiments indicate that observers experience difficulty excluding this associated contextual information. The next section will review key findings in the high-level attentional control literature consistent with this claim.

Attention and Temporal Limits in Perception

If observers do use contextual scene knowledge in the allocation of visual attention, this requires that scene knowledge must be available rapidly and with minimal cognitive investment. The following section reviews the evidence that these conditions are met.

Attention and conceptual short-term memory. For contextual features to guide the visual selection of objects, the mechanisms governing ACS would have to be relatively flexible. These control mechanisms would have to be

malleable in two respects. First, given that contextual associations are formed on the basis of an observer's experiences, attentional mechanisms would need to be able to gradually change over time to reflect these regularities. As mentioned, in much the same way that spatial attention reflects the laws of perceptual organization, higher order attentional systems are governed by higher order perceptual and conceptual knowledge. Second, observers would need to be able to control the activation of these contextual associations so that only task relevant associations are active at a time (Bar, 2004). A variety of research programs evaluating visual attention demonstrate a remarkable degree of flexibility in both of these senses. The following experiments are a sampling of recent findings in this exciting new research area.

The selection of stimuli according to abstract, conceptual criteria is predicted by the conceptual short-term memory hypothesis (Potter, 1976; Potter, 1993). Potter argues that fleeting conceptual representations of objects and features are ubiquitous in mental processes and emphasizes their role in fairly basic perceptual operations. Conceptual short-term memory (CSTM) putatively acts as an interface between perception and long-term memory, permitting volatile candidate conceptual representations to be compared with durable mnemonic traces. When an object is successfully associated with structures in long-term memory, a lasting representation is constructed that is available for free report and the guidance of behavior. The mechanisms of CSTM are implicated in wide variety of tasks including reading, object and scene perception. Critical to our current discussion, the activated concepts in CSTM are

integrated both with each other and associated items in LTM. In one demonstration of the heterogeneous nature of these memories, polysemous words temporarily activate multiple possible meanings before contextual constraints bias selection towards semantically consistent possibilities (Swinney, 1979). CSTM mechanisms are not restricted to the representation and sustenance of these temporary representations, but are hypothesized to play an active role in the construction of elaborate multidimensional durable mnemonic objects. In terms of psychological processes, CSTM is active between the identification of an object and its consolidation into visual short term memory. CSTM would be well suited to process information available in a single fixation.

The strongest evidence for CSTM is presented by Potter (1975, 1976). Participants viewed rapidly presented photographic scenes after being provided with a verbal label or picture of some target object. After viewing sequences of images at speeds of around 100 ms / item, subjects were able to quite reliably detect the target image in the sequence. Intriguingly, accuracies for either verbal labels or an actual preview of the target image were comparable (although later, a picture target advantage obtained), suggesting that whatever memory system was involved in the detection of the target image in the sequence relies on relatively abstract representations quite effectively. When subjects were not cued to the target category and simply given a recognition test regarding the same target image, accuracy was much lower.

This vulnerability of uncued, briefly viewed images to interference by preceding and succeeding images is referred to as conceptual masking.

Conceptual masking is distinct from perceptual masking because it involves interference during, as opposed to prior to, the identification stage (Intraub, 1984; Breitmeyer & Ogmen, 2000). The vulnerability of these memories was tested directly by Potter and colleagues (Potter, Staub, Rado, & O'Connor, 2002). Observers were presented with an RSVP stream followed by a recognition task, as in the previously discussed experiments. The nature and timing of the recognition test following the RSVP sequence was manipulated. When observers were required to wait several seconds between viewing the RSVP sequence and attempting to recognize individually presented items, observer performance declined as a function of the interval. As one might expect, performance also declined across the recognition test, such that accuracy was highest for items at the beginning of the list. Intriguingly, there was no recency effect in terms of the RSVP sequence. Items late in a given image stream were no more likely to be correctly recognized than those at the beginning. This supports prior work showing that once a picture is no longer being processed because a subsequent image is being viewed, it is not any more or less subject to interference as more and more images are processed (Potter, 1976). In a second series of experiment, the authors presented observers with the same RSVP recognition task (Potter, Staub, O'Connor, 2004). The conceptual relationship between the old items and the lures was manipulated in order to determine whether picture recognition accuracy is supported more by semantic or conceptual properties as opposed to the featural details of the memorized images. The hypothesized relationship between these multiple memory systems was as following. Both

CSTM and PSTM (pictorial short-term memory) are active on the order of several seconds. Over time, the detailed visual information, contained in PSTM, is lost, leaving only the CSTM information. The information in CSTM must either be matched with items in long-term memory and tokenized into a durable representation or it will be lost. Items that are tokenized successfully are available for report for longer periods of time. Observers were presented with both actual photographs and verbal labels during the recognition test portion of the experiment. The serial position of true and false items within the presented recognition lists was manipulated. The authors present comparable recognition performance with both verbal and photographic probes in many conditions, with an overall advantage for picture probes. While conceptually related lures were more likely to be falsely recognized, the strength of the effect diminished over the course of the recognition test, suggesting the activated CSTM representations faded over the course of testing.

The role of CSTM extends beyond picture perception, however, and may also support various types of verbal behavior (Potter, 1999). It has been hypothesized that the recall of sentences longer than an individual's short-term memory word span may involve temporarily active traces in long term memory. Within this account, as a listener hears a sentence words are entered to the standard modal phonological loop. At the same time, the meaning of these words are extracted via CSTM and selectively activate matching items in long-term memory. When the sentence is recalled, the activated lexical items in long term memory are more likely to be sampled because of their recruitment by CSTM.

Observers in this study were presented with individual words from sentences at rates up to 10 Hz, much faster than phonological encoding speeds. Potter argues that CSTM must be somehow distinct from the types of conceptual representations available in long term memory, because participants are able to associate a given conceptual representation to the current circumstances. Stated alternatively, the memory system must create a token for a given mnemonic object apart from the type information used to label the token. This token, which is quite similar to an object file, is argued to contain pointers information in LTM (both the type and its associates) as well as contextual and episodic information. Under the proper circumstances this CSTM representation may reach awareness and be available for subsequent report over some time interval.

In terms of our current discussion of semantic ACS, it has been hypothesized that CSTM may play a role in the guidance of visual attention (Belke et al., 2008). CSTM representations are available on a timescale that would permit their use in selecting fixation locations. The relative efficiency and minimal cognitive investment involved in the creation of these memories is another attractive feature of this system. The following experiments report recent findings regarding attentional capture by conceptual features.

Rapid access to affective information. Influential researchers argue that since selective attention is used to pare down the wash of data across the senses and emotional significance often marks biologically important data, it is likely the mind uses emotional significance to identify objects that ought to be attended (Lang, Greenwald, Bradley, & Hamm, 1993). The direction of attention

toward affective targets involves ACS just as much as any other attentive behavior. Generally speaking, emotional ACS involve control parameters and target representations over which the observer has little control. For example, participants have been shown to fixate attractive opposite-sex conspecifics in a manner largely independent of currently active goals or tasks (Duncan, Park, Faulkner, Schaller, Neuberg, & Kenrick, 2008). Here we have a predictable and complex attentive behavior (fixation) that occurs on the basis of a relatively sophisticated perceptual evaluation (attractiveness) outside of an individual's control. The rapid processing of emotional stimuli are another domain where the rapid access to high-level perceptual and conceptual information seems to proceed in a manner insensitive to manipulations that usually strain visual attention.

Emotional reactions to stimuli almost always involve the furthering or impeding of some biological goal (Arnold, 1960). In order for emotional tagging of stimuli to have any utility in rapid deployment of selective attention, emotional processing of affective stimuli would need to occur quite quickly (Compton, 2003). Evidence from electrophysiological studies indicate brain activity in the ventromedial prefrontal cortex 150ms after stimulus (spider image) onset (Carriete, Mercado, Tapia & Hinojosa, 2004). The ventromedial prefrontal cortex is believed to be involved in threat processing. In this study the threatening stimuli were masked and the participants had no awareness of the threatening stimuli. Psychophysiological studies which monitored biological indicators of threat detection (e.g. blood pressure, skin conductance, heart rate, corrugator

activity) found a similar rapid response, with reliable changes within 500 ms of stimulus onset (Codispoti, Bradley, & Lang, 2001). Codispoti and colleagues presented stimuli to participants for 500 ms and found similar patterns of emotionally linked physiological response as previous studies in which stimuli were presented for 6 s. It would seem as though biological preparedness for threats reaches asymptote quickly, remaining stable after the first 500 ms. Researchers argue that this indicates that stimuli continue to be processed even after presentation. It seems that not only the central nervous system, but the peripheral nervous system as well, can respond to emotional stimuli in well under one second. This window of time that would permit selective attention to utilize emotional significance as a source of information in situations that would require rapid responses.

Lesion studies involving bilateral simultaneous stimulation provide converging evidence that threat-related stimuli are preferentially processed (Vuilleumier & Schwartz, 2001). Two subjects with right parietal focal lesions demonstrated extinction of briefly presented stimuli in their left visual field. However, when images of spiders were presented in the left visual field, subjects were able to correctly identify images as accurately as controls. It should be noted that the spiders were matched with flowers in terms of low-level visual properties by rearranging the lines in the illustration.

Emotional salience engages attention. Codispoti, Bradley, & Lang (2001) presented participants with an abrupt auditory probe while they were presented with affective stimuli. The typical response to a 50 ms presentation of a 103 dB

tone is a startle response, which almost always entails a blink. By measuring blink suppression, researchers hoped to evaluate attentional involvement with the affective stimuli. Blinks were inhibited longer for emotionally valenced, either pleasant or unpleasant, stimuli. However, when subject did blink, the magnitude of the startle reflex was greater when participants were presented with negative, as opposed to positive or neutral, stimuli. Similar results obtained in a study by Cuthbert, Schupp, Bradley, McManis, & Lang (2001). Researchers concluded that affective information is used to modulate the startle reflex, leading to heightened startle reactions in the presence of negative, or threatening, stimuli. Other evidence for a strong relationship between emotion and attention can be found in a study by Anderson & Phelps (2001). Using a rapid serial visual presentation paradigm, researchers determined that the attentional blink is attenuated when the second target is emotionally salient. This attenuation was not evident in a participant with damage to the amygdala.

A recent study by Phelps, Ling, & Carrasco (2006) may further illuminate this relationship between attention and emotion. In an investigation of transient, covert attention, researchers presented participants with an orientation discrimination task using gabor patches of varying contrast. The patches could be primed by a fearful or a neutral face in the center of the screen. Participants had lower contrast thresholds when presented with the frightened, as opposed to the neutral face. In a second experiment participants were presented with a neutral or fearful face cue in either a peripheral location or distributed about the screen. The location of the peripheral cue changed across trials. Participants had

lower contrast thresholds with the frightened faces in both the peripheral and distributed conditions. Interestingly, these results show independent contributions of emotion and spatial attention, such that the peripheral cue, in the quadrant of the screen where the target was to appear, resulted in the lower contrast threshold than the distributed cue, likely because the distributed cue spread attention evenly about the screen. However, the distributed fearful cue still resulted in lower thresholds when compared to the distributed neutral cue. Researchers conclude that reciprocal projections from the amygdala, which processes threats preattentively, loop back to the early visual areas of the extrastriate cortex, increasing the speed and accuracy of visual processing. Additionally, while the effects of emotion on perception may come about in this experiment via the moderating influence of transient, covert attention, there is evidence that emotion may have a potentiating effect on visual processing even in the absence of attention. When the cue was distributed evenly across the screen, so there was no cue for covert attention to use to localize the target, there still were lower contrast thresholds. This study utilized fearful faces because fearful faces provide ambiguous information about the environment. The information about the environment is ambiguous in so far as it signals a threat, but does not identify it.

Evolutionary psychological arguments for the advantage that affective stimuli obtain are numerous. Theorists have identified a number of domains in which emotion may have influenced fitness, but of particular interest are the areas of attention, perception, and learning. Explicit criteria have been

formulated to define the boundaries within which an emotion can be accurately labeled as an adaptation (Tooby & Cosmides, 1990). For an emotion to be considered an adaptation, ancestral populations found themselves presented with a situation with great enough frequency as to constitute an "adaptive problem." This situation must be identifiable by situation-specific cues. Additionally, these cues must be monitored by algorithms that detect situations and then react in a manner that increases fitness. Threats in the environment were present in abundance and constituted an adaptive problem. Moreover, these dangerous situations can be sometimes be quickly detected utilizing visual cues.

While traditional explanations of perceptual advantages for affective stimuli relied on evolutionary arguments, recent evidence suggests that the categories of affective stimuli that these stimulus categories can be learned. Blanchette (2006) presented observers with a visual search task containing various affective targets. Observers were presented with threatening objects that might have been encountered in a humans' evolutionary environment of adaptation (e.g. snakes, spiders) as well as those that represent novel developments in material culture (e.g. syringes, guns). Subjects were instructed to indicate whether any of the 4 or 9 objects presented did not share category membership with the distractors. As one might expect, an advantage for negative stimuli obtained, such that when they were the discrepant items observers responded more rapidly. Critically for our current discussion, this effect was greater for artifactual threats than evolutionary threats. In a second

experiment, participants were presented with the same task and stimulus categories, but instead of presenting photographs, cartoons of the object categories were employed. Despite the lack of verisimilitude, observers again demonstrated an advantage for artifactual threatening objects over biological threats. These data suggest that the mechanisms involved with the detection and localization of threat related information in the environment are not as hard wired as initially posited.

In a recent methodologically novel demonstration of access to the semantic attributes of heavily masked French words, observers were tasked with the identification of neutral and emotional word stimuli (Gaillard, De Cul, Naccache, Vinckier, Cohen, & Dehaene, 2006). These experiments were designed to measure the effects of both familiarity and meaning. Familiarity was manipulated by presenting the words repeatedly with increasing or decreasing mask durations. This allowed the researchers to test identification performance on the same words with the same mask duration under conditions where subject either had or had not consciously perceived the word. As one might expect, familiarity increased the accuracy of word identification, allowing observers to report words with very short target mask asynchronies. Of greater importance for the current discussion was the result of the semantic manipulation. Observers reported emotionally charged words more often and more accurately than control words. This was the case both when observers had recently consciously seen the word (in the increasing masking condition) and when observers had not recently consciously seen the word. Great pains were taken by the researchers

to control all relevant word statistics (orthographic neighborhood density, frequency, etc.). In many cases, identification benefits accrued for emotional words which differed from control stimuli in only one letter (danger vs. ranger).

In all these experiments, we see participants able to rapidly make relatively abstract, conceptual characterizations of linguistic and photographic stimuli. This suggests that early attentional filtering mechanisms are capable of making sophisticated categorizations, relying on long-term knowledge about objects and the world.

Rapid processing of scene semantics. As mentioned previously, evidence that semantic factors can influence scene processing within a single fixation is presented by Gordon (2004, 2006). This prioritization of semantically inconsistent objects within one fixation is consonant with previous research indicating that semantically inconsistent are generally more likely to be attended. Hollingworth & Henderson (2000) found that subjects were more likely to detect changes in a change blindness paradigm when the changes were made to semantically informative objects. Objects were considered semantically informative when they were inconsistent with the schema for the scene. Individual semantically consistent objects are not considered informative because they all point to the same scene category. Further research demonstrated an inconsistent object advantage even when the influence of eye movements were controlled, suggesting that inconsistent objects are preferentially encoded. Within the memory schema hypothesis proposed, semantically consistent objects are set to their default schema value, but

inconsistent objects are stored more accurately. Regardless of these details regarding the processing of scene semantic properties and their relations, this finding is helpful in our discussion of semantic ACS because it indicates that the semantic properties of many illustrated scene objects can be rapidly extracted, a scene categorization can be made, and the location of the schema-inconsistent objects can be determined in as little as 150 ms.

In another demonstration of rapid, seeming pre-attentive processing of photographic scenes, Li, VanRullen, Koch, Perona (2005) conducted an investigation into rapid scene categorization while subjects simultaneously performed an unrelated task. Subjects were presented with sequences of letters centrally and were instructed to make same different categorizations. In the dual-task condition, images were presented simultaneously at varying degrees of eccentricity. Subjects were presented with scene categorization tasks consisting of facial gender, animal detection, and vehicle detection. Several control tasks using synthetic stimuli were also employed. Subjects performed comparably in both the single and dual task conditions, indicating little attention was required. In the control tasks, subjects exhibited generally poor performance, despite the introduction of a stronger, more redundant stimulus. Surprisingly, subjects were able to detect the presence of an animal in more than one image when two were presented simultaneously. One control condition showed performance similar to that in the scene categorization condition. When subjects were discriminating upright letters, because the stimuli are meaningful and familiar, subjects were able to categorize accurately regardless of dual-task load. The authors conclude

that meaningful objects with which subjects have had extensive experience may be processed outside of attention.

While in many studies of scene perception there is sufficient time for cognitive mechanisms that involve feedback driven hierarchical constraints, some still pose a major hurdle such an interactivist account. Research by Thorpe and colleagues demonstrates that subjects are able to reliably identify common categories of objects and scenes and generate the appropriate motor response in as little as 150 ms (Thorpe, 2002). In a series of studies observers were presented with photographs of animals for very short durations (~20ms) (Thorpe et al., 1996). Subjects were to indicate whether the photographic scene contained an animal in a go/no-go paradigm. Animals were drawn from a range of categories including mammals, fish, insects, and reptiles and presented at a range of scales. Distractor stimuli included the same sorts of natural contexts in which the animals would appear, however, these scenes contained no visible animals. Despite the brief exposure, subjects were quite accurate and responded rapidly. The distribution of responses were sorted in order to find the minimum reaction time, or the time at which correct responses significantly outnumbered incorrect responses. This minimum reaction time, including time to plan and execute a motor response, was 250 ms. This rapid identification of critical objects was extended in a finding involving vehicle detection (VanRullen & Thorpe, 2001), demonstrating that this form of rapid scene processing is not restricted to natural categories. Interestingly, familiarity does not seem to facilitate this type of object detection task (Fabre-Thorpe, et al., 2001). After

training with a subset of images over 14 days, observers demonstrated similar reaction times for familiar or novel scenes in rapid scene categorization task. This rapid scene processing can occur rapidly and in parallel across multiple images despite differences in viewing conditions, structural contexts, scale, and other inter-image variables. Observers can detect the presence of an animal in two scenes presented on either side of fixation as rapidly as a single scene and with comparable accuracy (Rousselet et al., 2002).

Because of the complications involved in plotting the time-course of these rapid scene classifications using only overt responses, convergent methods have been employed. Electrophysiological investigations demonstrate differential activity in the frontal lobe for trials with animal-containing scenes in as little as 120ms after stimulus onset (Thorpe et al., 1996). Tracking eye fixations represents another sensitive and highly ecological measure of scene processing. Observers were presented with two scenes for 20 ms offset 6° on either side of fixation and instructed to saccade toward the scene that contained an animal (Kirchner & Thorpe, 2006). The minimum saccadic reaction time, defined as the 10ms time bin that contained reliably more correct than incorrect responses, was 120ms. Surprisingly, if one assumes 25 ms is required to plan a saccade, the presence of an animal can be reliable detected in one of two briefly presented scenes in approximately 100 ms.

These results are indicative of a response that was generated on a single feed forward pass (Fabre-Thorpe et al., 2001). It should be noted that this is not an argument about whether top-down feedback plays any role in scene

perception, but rather a demonstration that many complex perceptual tasks can be accomplished in a purely feed-forward manner. This feed-forward position is supported by a variety of arguments. First, recent evidence indicates that even cells relatively early in the visual system are highly selective (Karklin & Lewicki, 2003). The non-linear behavior of these early visual cells stands in contrast to traditional filterbank accounts of early visual processing (Hubel & Wiesel, 1962, 1990). Additionally, some ultra-rapid classification studies have demonstrated that responses are just as fast with black and white images (Delorme et al. 2000). The authors argue that this is consistent with accounts in which information from the achromatic magnocellular pathway reaches the visual cortex before the chromatic information present in the parvocellular pathway (Nowak & Bullier, 1997). Additionally, neuroanatomically plausable feed-forward computational models have been able to perform a variety of scene processing tasks including face detection (Van Rullen et al., 1998) and animal detection (Serre et al., 2007).

The neuroanatomical constraints that drive this feed-forward argument are beginning to be well understood. As shown in Figure 5, each successive processing stage has been characterized to some extent.

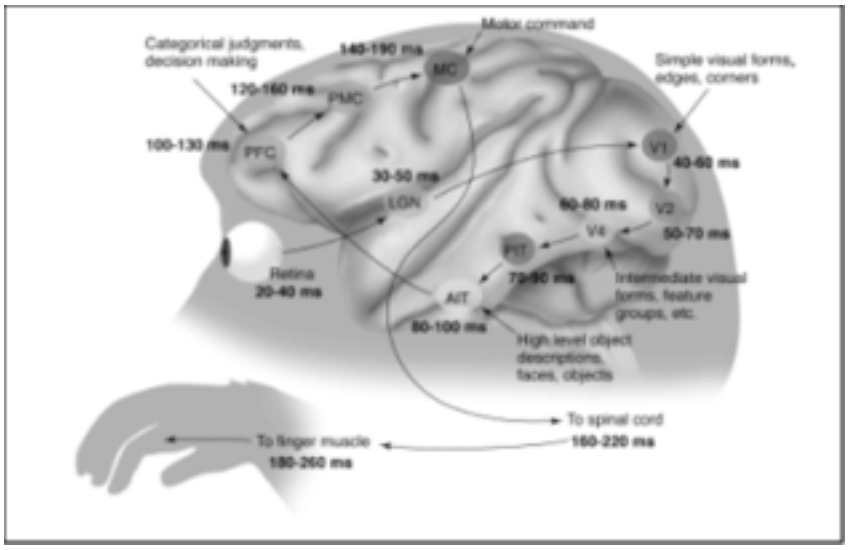

Figure 5. A schematic temporal characterization of processing in a rapid scene classification task (Image reproduced from Thorpe, 2002)

It should be noted that this temporal analysis does not assume that processing occurs serially or address whether these processes cascade between stages. These latencies represent the earliest reliable activity at a given stage of visual processing in response to a stimulus, not the termination or resolution of those processes. Latencies between the retinal surface and the retinal ganglion typically run around 20 ms (Sestokas et al., 1987). Schmolesky et al. (1998) measured firing latencies across the macaque visual system. Their results are particularly useful because all timing observations were collected in a single lab and all the animals were prepared using similar methods. The subjects were anesthetized and presented with synthetic stimuli while single cell recordings were collected. Spike trains from the retina reach the lateral geniculate nucleus in around 30 ms, in the case of the magnocellular layers, and 50 ms, along the parvocellular pathway. Information arrives at V1 between 50 and 70 ms following

stimulus onset, depending on the laminar layer. Neural firing in response to a visually presented stimulus has been demonstrated in V4 in 60 to 80 ms after stimulus onset. Following the elaborate processing that occurs in V4, object identification (e.g. animal) likely occurs in the inferotemporal cortex (Tanaka, 2002). Signals would require an additional 40 to 20 ms to reach this high-level visual recognition area. Once a candidate object is identified, a decision, presumably involving prefrontal areas would be required. Once a decision is made, pre-motor and motor areas must represent and execute the appropriate response with additional time required for transmission of the efferent motor signal to the spinal cord and then the hand. Given the timing estimates described above, little time is left for feedback from a higher-processing stage to a lower one.

It should be noted that the sequence of stages described above does not include the possibility of subcortical processing. Neuroanatomical studies indicate that older, direct connections, such as those from the thalamus to the amygdala, may provide coarse information without the elaboration that is believed to occur in the inferior temporal cortex (Fendrich et al., 2001). It seems reasonable that certain types of affective stimulus processing could proceed without the elaboration enabled by the recruitment of higher, cortical visual areas. However, the range of classifications that can be rapidly performed in the previously described series of experiments (e.g. animals, vehicles, scene categories), are difficult to account for in terms of a entirely or mostly sub-cortical mechanism.

In the studies conducted by Thorpe and colleagues, observers were presented with a single object detection task. As mentioned, detection accuracy was quite high and response times were surprisingly low. Fei-fei et al. (2005) conducted an investigation into this rapid scene categorization while subjects simultaneously performed an unrelated task. Subjects were presented with sequences of letters centrally and were instructed to make same different categorizations. In the dual-task condition, images were presented simultaneously at varying degrees of eccentricity. Subjects were presented with scene categorization tasks consisting of facial gender, animal detection, and vehicle detection. Several control tasks using synthetic stimuli were also employed. Subjects performed comparably in both the single and dual task conditions, indicating little attention was required. In the control tasks, subjects exhibited generally poor performance, despite the introduction of a stronger, more redundant stimulus. Surprisingly, subjects were able to detect the presence of an animal in more than one image when two are simultaneously presented. One control condition showed performance similar to that in the scene categorization condition. When subjects were discriminating upright letters, because the stimuli are meaningful and familiar, subjects were able to categorize accurately regardless of dual-task load. The authors conclude that meaningful objects with which subjects have had extensive experience may be processed outside of attention.

Taken together, these studies indicate that observers are able to rapidly, and efficiently, extract semantic information from scenes. While the specific

mechanisms supporting this performance are uncertain, scene knowledge available so rapidly could be used to direct visual attention towards items relevant to the current goal of visual attention.

Attention to meaning. We have reviewed the evidence that scene knowledge is available rapidly in scene viewing. The following findings review evidence that the attentional mechanisms are sensitive to semantic properties of briefly presented visual stimuli. In many of these cases, an attentional blink (AB), or deficit for targets appearing close in time, is the measure of attentional engagement. A more detailed description of the AB follows this sampling of recent evidence for semantic, or conceptual, ACS.

Evidence for a general semantic AB following the capture of visual attention is presented by Maki & Mebane (2006). Observers were presented with target words in a false font RSVP paradigm. Participants reported target words presented in black in an RSVP stream at the end of each trial. A variety of distractors sharing different levels of similarity to the target were embedded in the RSVP sequence. In contrast to the false font stimuli, some trials contained critical distractors consisting of colored words or consonant strings that preceded target items at certain set lags. Observers were less likely to accurately report target strings when they were preceded by these attention capturing distractors. Specifically, the semantic characteristics of the distractors drove the effects, with those stimuli that were most word-like resulting in the greatest costs. These results demonstrate that searching for any meaningful item among mostly

meaningless distractors can result in an AB following a meaningful but visually distinct distractor.

Recent evidence suggests that an observer's current emotional state influences attentional capture by critical distractors in an RSVP sequence (Most et al., 2010). Heterosexual couples were placed in an experimental setting wherein female participants were led to believe that their partners were completing an attractiveness rating task involving either landscapes or women. While male participants completed these ratings tasks, female participants completed an RSVP task where they were to indicate whether or not a rotated landscape was presented on a given trial. Critical distractors were selected from an affective image database and included negative, arousing images. Those female observers who rated their unease with the attractiveness rating task demonstrated a larger AB for negative images. These findings suggest that semantic picture processing is modulated by the observer's emotional construal of the experimental context.

In another AB demonstration of semantic ACS, observers were presented with an RSVP task in which they were to selectively encode and report the identity of words on the basis of high-level semantic features (Barnard, Scott, Taylor, May, & Knightley, 2004). Observers were presented with lists of 35 words at 110 ms/item and instructed to recall only those words that referred to professions (e.g. baker). Despite the arbitrary and rather abstract criteria used to define targets in the sequence, observers had little difficulty accurately reporting those words which belonged to the appropriate category. However, report

accuracy depended heavily on the semantic properties of a distractor word that preceded the target at various lags. Most words in the sequence referred to natural objects and location (e.g. archipelago, cloud, thicket). Critical distractors were presented in two conditions, each with a unique level of semantic relationship between the distracting word and the target-defining category. Distracting words with low levels of semantic relatedness to the target category included various common household objects (e.g. telephone, couch). Distractors with high levels of semantic relatedness described human roles other than professions (e.g. father, tourist). The temporal relationship between these distracting words and targets on a given trial was manipulated. Observers were less likely to detect target words when they were preceded by distracting words of high semantic relatedness. The costs associated with the detection of a related distractor were distributed in a manner similar to the classic AB effect (lag-1 sparing, gradual recovery of performance over the 500 ms). This result is interesting because observers were not required to make any overt response to the related distractors.

It would appear as though the types of semantic features used to selectively encode and retrieve items compatible with current ACS are rather coarse and do not have detailed denotative meanings. The magnitude of the blink effect associated with a particular type of distractor was predicted by measures of semantic related using latent semantic analysis (LSA) (Dumais, 2004; Landauer, Foltz, Laham, 1997). LSA uses measures of the co-occurrence of words in large corpora of text to quantify their contextual-usage meaning.

There are four steps to an LSA treatment of corpora. First a “bag of words” representation is generated describing the frequency for each word in some larger unit of words (sentences, paragraphs, etc.). In the second step, this table is then transformed to normalize the frequencies relative to upweight infrequently occurring words and downweight frequently occurring words. Next, these normalized frequency tables are then decomposed into an arbitrary number of semantic dimensions that describe the frequencies in terms of an arbitrarily defined number of hidden dimensions. Researchers are able to specify the number of dimensions used to describe the distribution of words. Increasing the number of dimension improves the fidelity of the quantitative semantic description of the target word. At the same time that more dimensions results in higher accuracy, diminishing returns result from the use of too many factors. Lastly, the similarity between each word and all other words is calculated in this new reduced multidimensional space. Similarity between vectors containing values along each of the inferred dimensions is quantified by the cosine of the two vectors (essentially a measure of the angular similarity between the two vectors in this high dimensional semantic space).

The results of Barnard and colleagues (2004) have been successfully described using a computational model. Barnard & Bowman (2003) argue that two semantic memory systems, an implicational system and propositional system, support regular semantic processing of stimuli. The implicational system represents what the authors refer to as “generic level” meaning, including connotation and category relations. This implicational system is capable of

rapidly extracting general semantic information and assessing the salience of the information in the context of the ongoing task. The salience of a given distractor in the current task is determined by both transient and enduring attentional dispositions. For example, emotionally charged or personal items might pass the implicational test despite being quite independent of current overt behavioral goals. The implicational system takes "the immediate products of visual, auditory, and body-state patterns." This information is considered unrefined and relatively direct.

Ariga & Yokosawa (2008) present additional evidence of selective processing of objects on the basis of abstract, non-visual ACS in an RSVP experiment using something like a modified Stroop involving kanji. Subjects were instructed to report the identity of a uniquely colored target in that sequence. An attentional blink, or temporary performance deficit following a critical distractor, was observed when a target character was preceded by a distractor whose meaning matched the target color. That is, if observers were looking for a character that appeared in blue, they would be more likely to miss this character following a character that means blue. This attentional capture effect indicates that the ACS employed by the observers were abstract enough that capture obtains for stimuli that share only semantic properties with the target. This is not to say that ACS have no detailed visual character. Rather, there is a not insignificant semantic component that can influence performance in tasks that do not require semantic processing of stimuli.

Koivisto & Revonsuo (2007) extend these findings, presenting evidence that task-driven semantic ACS can influence the likelihood of detection in inattentional blindness tasks. Observers were instructed to rapidly encode and remember a collection of 4 simultaneously presented words or line drawings. After completing several expected trials, on a critical trial observers were presented with an unexpected word (in the picture condition) or picture (in the word condition) at fixation. As expected, inattentional blindness was observed for unexpected words or objects. Importantly, this inattentional blindness interacted with the semantic properties of the unexpected stimulus, such that when the word or image matched the semantic category of the stimulus in the primary task (either animals or furniture), inattentional blindness was less likely.

In a study evaluating the role of long-term object representations in a visual search task, observers were presented greyscale photographic images by Olivers (2010) and instructed to locate a traffic sign with an associated color (e.g. stop sign). Despite the fact that the target was always a greyscale image, observers were slower to locate and respond to the target on trials where there was a red distractor. Olivers concludes that attentive behavior is guided by information stored in long-term memory representations and that the use of these representations is automatic. That is, observers could not exclude color knowledge regarding the stop sign from the attentional filter they employed despite the face that this harmed their performance in the task.

Moores, Laiti, & Chelazzi (2003) present compelling evidence that visual attention is sensitive to the associative relationships between objects when

observers are conducting visual searches. Observers completed a visual search task containing photographs of common objects. Search displays were presented very briefly followed by a patterned mask. Participants searched the displays for a verbally labeled target object. Critically, on some trials these target objects appeared along with an associated item. These associations were varied including tool-object (hammer-nail), resource-product (cow-milk) and conceptual relationships (statue of liberty-american flag). Free recall and recency judgments regarding distractors demonstrated that objects associated with a search target are preferentially processed. These direct measures of distractor processing may have given participants an incentive to attend to items other than the target. In Experiment 4, observers simply completed the primary task of determining whether a target object was present in the display. Here observers had no reason to attend to items other than the target. Here the presence of associated distractors reduced accuracy and increased latency, but only on target absent trials. It appears as though associates were transiently treated as targets, resulting in higher false alarms (18% vs. 10%) and slower responses on target absent trials. The authors argue this is due to the strong advantage for target objects in the competition of the target with its associates. Also of interest is the lack of a spatial attentional effect in Experiment 3. In this experiment, there was a red dot probe that appeared on one of the objects in the visual search array after some random interval. On target absent trials, this red dot could appear on either associatively related or unrelated objects. Subjects were no faster to respond to the onset of the dot when it appeared on the related objects. This is

consistent with selection of associated objects occurring at a later, non-spatial stage of perceptual processing.

Figure 6. Examples of the associative pairs employed in Moores, Laiti, & Chelazzi (2003) (Fig 2, p.184)

While the results of these experiments are provocative and have clear consequences for the current set of experiments, there are three key issues that undermine this demonstration of associative attention. First, observers were completing a detection task. The use of a detection task has two consequences. Because observers completed the coarsest of perceptual determinations, detection, the experiment provides little information about the specific type of information the observers used to complete the task. Observers could be identifying the target based on detailed or coarse object knowledge. In a related result of the design, the ACS required for observers to detect an item that may or may not be present is in many ways unlike the attentional set required to selectively encode a target. In some circumstances observers are searching for an object and must both identify the object and perform some perceptual

operation on the object. For example, an observer might be required to answer a question about an object's pose or color. It is not clear what influence depictions of associated concepts will have in these circumstances. Second, observers were encouraged to respond quickly as response time was a dependent measure. Theories of attentional control are more precisely tested under unspeeded conditions (Leber, 2004; Norman & Bobrow, 1975). If selective mechanisms are to be strained in isolation from post-selection operations, it is important to use accuracy as a dependent measure in an unspeeded discrimination. For example, if a speeded discrimination is made, observers' performance may reflect response conflict (Gratton et al., 1988). While these post-perceptual effects are important, measuring them together with perceptual effects can make it difficult to evaluate claims carefully. Third, because the target and the associated distractor shared a common onset, little can be said about the timecourse of object-associate interference. For example, Auckland et al. (2007) found no effect of associated items when they shared a common onset with the target in an unspeeded forced-choice discrimination. Unlike the work by Auckland and colleagues, participants in these experiments had a particular target in mind. The effects of context could be quite fast acting, given that the participant already has a target in mind, but this is unclear in this design. Each of these three methodological concerns is addressed in the present experiments using a design sensitive enough to detect contextual costs on target-present trials.

These studies demonstrate the specific types of flexibility we described as necessary if observers were to selectively attend to associated contexts when searching for a categorical target. Observers attend to both target objects and information associated with a target object. This means that ACS are controlled by a mechanism that reflects the redundant structure of perceptual events. Further, these associative structures are selectively deployed on a trial by trial basis, such that observers preferentially attend to items relevant to the current trial primarily. The following section will formulate the motivating theory for these experiments more explicitly.

Attentional capture and control. In order to evaluate theories regarding the selection criteria utilized by attentional mechanisms, and the degree to which observers can choose these criteria, one requires a paradigm that can sensitively measure these systems. One particularly well replicated failure of temporal visual attention involves difficulty detecting targets presented in rapid succession. The current experiments use an attentional blink as the measure of attentional capture by the related context. The attentional blink, or a temporary inability to process targets presented in rapid succession, was first demonstrated by Raymond, Shapiro, & Arnell (1992). Observers were presented with an RSVP sequence of containing target letters and distractors. Observers were instructed to detect and identify multiple targets presented with various onset asynchronies. Observers were less likely to detect subsequent targets following a detected prior target. The costs associated with the encoding of an initial target begins approximately 100 ms after target offset and persist for another 400 ms. This

decreased likelihood of detection is known as the attentional blink (AB). Critically, targets that are presented immediately following the initial target show no costs. This effect, known as lag-1 sparing, has been central in various accounts of these costs.

Raymond and colleagues initially argued that the attentional blink was due to inhibitory mechanisms that exclude irrelevant information while the initial target is being processed. The elaboration of a visual representation within VSTM may be vulnerable during the first few moments of encoding. By shielding these elaborative processes from competing visual information, the first target is made available for report but later targets are missed. According to this protective account of the AB, the degree of inhibition should vary as a function of the difficulty of perceptual processing required to individuate and identify the first target. On trials where the first target was easy to process, little or no attentional blink should be observed because the processing of the first target would be interfered with by successive targets to a minimal extent. Contrastingly, on trials requiring sophisticated processing of the first target, observers will be more likely to miss successive targets. Lag-1 sparing occurs because two targets presented in immediate succession are both represented in VSTM simultaneously, because they map to similar responses they interfere with one another minimally.

This early selection account of the AB was falsified when Shapiro, Raymond, & Arnell (1994) presented observers with RSVP sequences containing targets that were either perceptually hard or easy (as assessed via a separate series of experiments) and observed a consistent AB effect. Instead, the authors

proposed a late selection account wherein both the initial and subsequent targets are processed to an extent where they are both identified as targets, but these temporary representations of the target category interfere in VSTM. This interference leaves only one of the two tokens available for report. A relatively late bottleneck has received significant empirical support. For example, Luck, Vogel, & Shapiro (1996) presented observers with RSVP sequences of words. The semantic relationships between words in the sequence was manipulated while electroencephalographic measures were gathered. The researchers monitored for an N400 pattern of activity which associated with violations of semantic expectancy. On some trials, observers were presented with words that violated semantic expectations. On other trials, words were appropriate for the context. Despite the fact that observers failed to report the target words when presented during the blink, the magnitude of the N400 wave remained relatively constant across numerous lag settings. This is consistent with observers processing the meaning of the stimulus, but failing to construct a representation that was available for free report.

Chun and Potter (1995) present a two-stage account of the attentional blink where the consolidation of a first stage representation prevents the consolidation of a successively presented competing target. When observers initially view a target in an RSVP sequence, the authors suggest that this triggers a volatile conceptual representation of the target object. This conceptual representation os hypothesized to include the target identity, it's membership in the target category, and similar semantic information. Before this initial

representation is available for report, the observer must consolidate this representation into a more durable form. This process takes some period of time. If the duration of this consolidation process extends beyond the point where a high quality representation of the second target is still available in the first stage, observers will be less likely to report the second target. Lag 1 sparing occurs because of temporal imprecision in the gating mechanism to the consolidation process. If the second target follows closely enough upon the heels of the first, then both targets are sent into the consolidation mechanism and both are successively tokenized.

In order for observers to be able to report the presence of a target in a sequence of rapidly presented items, they must construct of a representation of the target category or type that is associated with this particular context. The process of generating this context specific and temporary representation of the target class is known as tokenization. Failures of tokenization are well characterized within the repetition blindness (RB) literature. Repetition blindness involves the failure to detect repeated target stimuli when presented in rapid temporal succession (Kanwisher, 1987). It is rather similar to the AB and the psychological refractory period (PRP). In each of these cases, stimuli presented in rapid succession suffers when temporal limits of human perceptual, cognitive, and motor capabilities are exceeded. The rapid post-categorical memory system that seems to be responsible for RB is hypothesized to maintain an innate bias against creating multiple tokens of a single object. When objects are presented in rapid succession, this conservative mechanism either fails to individuate a

second token or merges the first token with the second. RB is unlike the AB in the following ways. First, subjects are typically instructed to freely recall the presented objects in RB studies (but see Kanwisher, Kim, & Wickens, 1996) whereas AB studies typically involve detection measures. Second, RB can occur when items are presented simultaneously, or in rapid succession (Kanwisher & Potter, 1990). In AB studies, the relative timing of the initial and subsequent targets is critical, as demonstrated by the lag 1 sparing phenomenon. A variety of other factors, including the discriminability of targets compared with distractors, the episodic distinctiveness of the individual targets, and the similarity between successively presented targets has been shown to influence AB and RB differently. While the attentional blink is clearly a different phenomenon the two may both result from failures of token individuation.

Both the Chun and Potter two stage model (1995) and the late-selection interference based account presented by Raymond and colleagues share hypotheses about the presence of a capacity limited stage where transient but reasonably elaborated stimuli compete for scarce resources. In the case of the interference based account, tokens compete in VSTM in a manner that is biased toward the first target that began the consolidation process. In the two stage model, representations in CSTM fail to be adequately tokenized because consolidation processes are preoccupied.

DiLollo, Kawahar, & Ghorashi (2006) and Olivers (2009) present accounts wherein the control of visual attention is central to understanding the attentional blink. Within both accounts, a guiding target representation is rapidly compared

with the successively presented visual stimuli. This representation is maintained within a relatively active attentional control system that is rapidly conducting sequential comparisons. These two theories diverge, however, in terms of how they address the AB from here.

DiLollo et al. (2006) argue that once a stimulus is encountered that matches, the attentional control processes that govern the maintenance of the target template cease with the presentation of the first target. In the absence of these signals, observers are able to identify an immediately following target because no other images, be they attended or not, have been presented. As soon as intervening distractor is encountered subsequent utilization of this attenuated target template is disrupted. The temporary loss of control model advanced by DiLollo and colleagues has been criticized for allowing the influence of capacity limitations similar to the Raymond and Potter models (Olivers, 2009). In this case, the capacity limitations are associated with the attentional control system and not the processes used to consolidate fleeting representations.

Olivers (2009) presents a novel computational model of the attentional blink that does not make reference to any capacity limited consolidation process. In this account, items are initially processed in a perceptual memory system where both low- and high-level information about the target is available. The gating mechanism that permits items to enter working memory is governed by an attentional control mechanism that maintains an attentional set for the target item. Critically, there is some temporal lag in the gating mechanism such that activation or inhibition of a matching item occurs approximately 100 ms after the

item is compared with the attentional template. When observers are presented with an initial target, this is quickly followed by an activation signal that boosts the perceptual processing of the next item. On trials where the initial target is followed by a second target, the second target is successfully consolidated. On trials where the target is followed by a distractors, this gating mechanism detects a strong mismatch (strong because of the activating signal generated in the presence of the first target). This strong mismatch results in an inhibitory signal being sent to the perceptual processing stage. On trials with target-distractor-target sequences, this inhibition falls on the second target.

Given the similarity of these various accounts it can be difficult to design experiments to arbitrate among them. However, all these accounts give a key role to the ACS that a subject is maintaining as they monitor the sequence of images. Regardless of the particular mechanics of the AB, in all cases the failure to identify subsequent targets occurs because multiple targets satisfy the target-defining criteria. This will be important when we consider the capture effects in the present experiments.

Chapter 3: Theory and General Methods

Attentional Capture in Object Search by Associated Contexts

Motivating theory. In our everyday lives we regularly visually explore our environment for objects based on incomplete information. We might be looking for the remote control, a set of car keys, or a missing pet. In cases where the object is familiar to us, this task is simplified somewhat as we have a general sense of the visual properties of our target. In the case of less familiar objects, our ability to quickly identify or locate specific categories of objects is harder to understand. One strategy that might be used to locate a known object category involves biasing attention towards associated contextual information. Observers can use knowledge about contexts where the target object is typically encountered as an additional cue to locate this categorically defined object.

The research literature reviewed is clear that: 1) object-context associations play a key role in object and scene processing, 2) scene schema and other high level scene properties are available quickly, 3) ACS mechanisms are informed by knowledge about objects, and 4) observers can establish ACS that match the current task at relatively abstract level. Given these facts, it is reasonable to hypothesize that associations between an object and scene can structure attentive behavior to support the localization and identification of task

relevant objects. The selection of a target object in a scene, particularly when the visual details of the target are uncertain, is not a computationally trivial task. If attentional mechanisms could be leveraged to isolate just those objects and features relevant to the object search, this would reduce the demands on the observer.

As reviewed previously, the influence of depicted context on the recognition of a target item is powerful (Palmer, 1975; Hollingworth & Henderson, 1999; Auckland, Cave, Donnelly, 2007). This influence obtains as differences in sensitivity and bias. The current experiments describe a contextual influence fundamentally unlike previous demonstrations. Typically, observers are presented with a target item and then must indicate retrospectively what the target or targets were. Observers are not looking for a particular object known in advance. The detection or identification of a prespecified target has been shown to be much easier than the recollection of unspecified items in a list (Potter, 1975). If observers show the influence of contextual scenes when asked to recognize a category of target specified on a particular trial, this will be a novel demonstration of contextual influence and will speak to issues involved in object recognition and context more generally.

In the present experiments, object-context associations are hypothesized to structure object search in the following manner. When observers wish to find a target object, anticipatory representations inform their overt and covert visual exploration of the environment. Once a set of ACS are active, currently viewed stimuli are compared with these guiding representations. The guiding influence

of contextual associations have been demonstrated in visual search paradigms that manipulate observers' familiarity with configurations of search arrays. In contextual cueing, observers are able to quickly locate targets in repeated displays because the configuration of the display acts as an attentional cue that can guide attention towards a target item (Chun & Jiang, 1998; but see Kunar et al., 2007). Learned contextual associations can result in costs if they invalidly cue the target location (Manginelli & Pollman, 2009; Fiske & Sanocki, 2010). It seems reasonable to predict something analogous would happen on the basis of pre-experimental object-context covariation. These experiments advance the discussion because here scene context is treated at a more abstract level (e.g. there is no licensed spatial relationship between the scene and target). Or stated another way, the influence of scene knowledge is measured primarily in terms of conceptual scene knowledge rather than spatial scene knowledge. Context in this case is treated as a relatively abstracted scene schema, rather than a particular configuration of searched items.

Observers searching for objects maintain a diffuse attentional set, with ACS informed by both target identifying and context identifying features. These attentional parameters may be especially broad when observers are searching for an object on the basis of a categorical description. The contextual representation employed in these ACS is a schematized description of the scene context where an object is typically encountered. When observers are presented with visual information that matches this contextual representation, scarce perceptual resources are allocated on the basis of this match. In typical scene

viewing, this is adaptive because, to the extent that object-scene associations represent the co-occurrence of scene features veridically, perceptual complexity can be reduced by the exclusion of irrelevant regions and enhancement of relevant regions.

Testing attentional capture. Researchers have spent a great deal of time considering what criteria need to be met for a stimulus to be said to truly capture visual attention (Logan, 1992). Only when there is no incentive in an experiment for attending to a stimulus can it truly be said to capture attention (Theeuwes, Kramer, Hahn, Irwin, & Zelinsky, 1999). As implied by the use of the term "capture", we seek to measure the direction of the observer's perceptual resources towards some object in violation of the observer's will. Typically, this is done by manipulating the salience of a distractor in some perceptually demanding task. For example, observers might search for a color target in the presence of an irrelevant onset or the reverse (Theeuwes, 1994). If observers attend to these salient distractors, despite the lack of an incentive to do so, the distractor can be said to capture attention and will result in performance costs ins the primary task. In the following experiments the task relevance of a distractor will be manipulated.

In order to put this theory to a strong test, contextual information and object recognition information will be manipulated independently and set in competition with each other. In this way we can measure the involuntary processing of contextual information while the participant is engaged in an object search task. Similar logic supports attentional interference paradigms such as

stroop (MacLeod, 1991) and flanker (Eriksen & Eriksen, 1994) tasks. In these cases there are competing information sources presented simultaneously in a common region of space. When distractor information bears certain relationships with the target, this interferes with processing in the primary task. Rather than the automatic and over-learned processing of irrelevant letter or word information in a letter or color identification task, the following experiments measure the obligatory processing of contextual information in an object search task. However, unlike typical instances of either of these interference tasks, the costs of this automatic processing in the following experiments will be measured in accuracy rather than response time for reasons described earlier. The way in which the objects and contexts are sequenced in these experiments is artificial but necessary to test determine whether participants must attend to the contextual image.

In the following experiments, observers will be presented with a series of object recognition tasks where no useful information is contained in a contextual image distractor on any trial. Despite this, the contextual images will capture attention because of relationships with the target item on a particular trial. This demonstration of attentional capture is contingently automatic in much the the same way as the demonstrations by Folk, Remington, & Johnston (1992). In these experiments, observers were presented with a cued visual search task. Targets appeared randomly at one of four locations. These targets were preceded by valid or invalid spatial cues. When the cues contained the target defining feature, they captured attention even when this impaired observers'

performance. When cues did not contain the target defining feature, observers were able to effectively ignore the cues. Observers were able to select a given attentional set, say for the color "red", but could not control the way in which this attentional set was implemented during selective processing. The observers could choose to selectively attend to red targets but could not ignore the red spatial cues that preceded the target containing display despite the fact that the form of the cues was visually distinct from that of the targets. In the following experiments, observers maintained an attentional set for a target object category. If this attentional set contains information about associated contexts, observers will be forced to attend these contexts even when it harms their performance in the object search task. In order to test this claim, it will be important to present the object and context in competition with each other.

Testing capture by associated context. These experiments test the claim that when observers search their environment for common objects, they do so with ACS that include information about schematized spatial contexts where objects are typically encountered. Perceptual processing of these associated contexts will be facilitated because the ACS bias processing towards not only target objects but also these associated contexts. Testing this prediction is difficult for several reasons. First, as indicated by Moores et al. (2004), targets are generally strong competitors with their associates. Designing an experiment powerful enough to detect an effect of associated distractor processing on target present trials is quite difficult (which is why RT measures are typically employed). In the current experiments, rapid serial visual presentation (RSVP) was used to

place targets and contexts in competition. Second, presenting observers with mixtures of target present and target absent trials can introduce complexities when interpreting data for reasons described earlier. The following experiments include both detection and discrimination designs, showing costs attributable to associated contexts in these two related perceptual tasks. Third, observers are sensitive to demand characteristics in studies such as this. If one claims that observers are involuntarily engaged by and process associated contextual information, it is important to structure the design so that doing so affords no advantage in the primary object search task. In these experiments, no contextual image contains the target object. Moreover, the contextual images are visually distinct from our targets. Fourth, selecting stimuli to test such a prediction is complicated on several levels. A broad sample of associated objects and scenes must be gathered. For each pair, multiple photographs of prototypical category members in a discriminable pose and scale must be collected. More subtly, appropriate control stimuli must be selected. Previous research in object recognition and contextual associations was compromised through inappropriate use of control stimuli (Biederman, 1981, Hollingworth & Henderson, 1998). In the following studies care was taken in stimulus selection. A broad range of object categories (69 or 71) were employed, ensuring the generality of the effect. A large number of the object-context pairs were selected from word association norms (Nelson, McEvoy, & Schreiber, 2004), and all pairs were independently rated by two observers with the lowest rated associations excluded. While each object category was encountered multiple times in all experiments, in most of the

following experiments an object image served as a target or lure only once. This prevented observers from using strategies based on low-level image properties. More importantly, a variable mapping design was employed such that the target and distractor on a given trial were chosen from the same pool of images. This ensured that target and distractor images did not differ in image statistics or novelty. The following experiments address each of these four concerns effectively.

General Methods

Presentation method. Because object recognition is an enduring research domain, a wide variety of experimental paradigms are available. In the current series of experiments, RSVP was chosen as a display method for several reasons. First, because we wished to know whether observers' attention is captured by associated contexts, observers had to be engaged in a task where selective attention is required to perform well (Leber, 2004). Observers who do not attend to a particular item in an RSVP sequence show poor memory for that item, ensuring a sensitive measure of observers' attention (Potter, 1976). Second, human object recognition performance is generally high, so images must be degraded to pull performance away from ceiling. In RSVP, images are masked, both by preceding and following images, across numerous dimensions. This makes the task challenging enough that manipulations have a chance to actually influence performance. Third, RSVP allows precise control over stimulus order, duration, and retinal position. To evaluate what sorts of images capture attention when an observer is searching for an object, a method must control, to

the extent that it is possible, observer gaze. RSVP stimuli are presented at or near fixation so performance is not limited by peripheral acuity. Further, because the stimuli are presented for brief intervals, there is no time for observers to make eye movements. While some subjects may make inadvertent eye-movements during the approximately one second trial, these deviations are cancelled out by averaging across trials and subjects. While clearly artificial in many respects, RSVP approximates natural scene viewing in others. The anisotropic distribution of photoreceptors along the retina surface, among other factors, forces observers to serialize their samples of visual information. In typical scene viewing, these fixations last only several hundred milliseconds before retinal input is suppressed and saccades are initiated towards another location. RSVP was initially designed to approximate the rapidly changing retinal input that accompanies the visual exploration of a scene (Potter & Levy, 1969). The selection mechanisms involved in sampling task relevant information at fixation from an RSVP stream may well be the same mechanisms that select saccadic targets and fixation duration during free scene viewing. Lastly, presenting the objects in isolation against a high contrast background ensures that the object outline is visible. The external outline of an object is particularly important in object recognition (Hayward & Tarr, 1997) providing information about part boundaries that can be used to identify objects (Hoffman, 1984; but see Sanocki, Bowyer, Heath, & Sarkar, 1998).

Stimulus selection. In the following five experiments, observers were presented with images sampled from a collection of associated object and

context photographs. Many of these object-context image pairs were generated using word association norms (Nelson, McEvoy & Schreiber, 2004). Candidate object-context associations were generated by choosing only those associations in the database with nouns as both cues and targets (e.g. boat-paddle). Only items with both forward and backward cue-to-target strength of at least .10 were chosen. Because these are free association norms, a forward cue-to-target strength of .10 indicates that 10% of subjects will report the target as the first word that comes to mind when asked to freely associate some word in memory with the target. Backwards strength indicates the likelihood a given cue word will be generated when the target is instead presented a cue (e.g. paddle-boat). These normed stimulus pairs were then supplemented with additional items as is common (e.g. Most & Junge, 2008).

There were several considerations that went into the selection of object-context pairs. First, in no cases were targets to be categories of humans (e.g. firemen) because observers demonstrate specialized capabilities for detecting the human form and face in photographs. This makes the interpretation of human detection performance problematic (see Mack & Palmeri, 2010 for a recent discussion). It should be noted here that many of the contextual distractors did contain photographs of humans. Because the variable mapping design of the study ensures that both associated and unassociated contextual distractors will contain human forms, differences in performance as a function of distractor relatedness will not be attributable to human images in the distractors. Secondly, target objects were chosen to represent a varied set of familiar object

categories. As can be seen in Table 1, the range of categories employed is quite broad. In order to ensure the generality of any findings, care was taken to include objects across a range of spatial scales and animacy categories. Appendix 1 contains all the pairs along with the associated images. Despite this category variability, within each category objects were selected that were typical tokens. This served two purposes. It ensured that subjects will likely be able to use previous encounters with the target category to establish a search template (Vickery, King, & Jiang, 2005; Bravo & Farid, 2009). If observers are to detect these objects on the basis of verbal label only, these objects need to have existing long-term memory representations that can inform ACS. More importantly, using familiar objects increases both the likelihood that the object will have associations and the likelihood that these associations will be shared across individuals. Once non-human, familiar target items were selected, the pool of possible associations was still further refined by selecting only those items for which multiple distinct, yet visually similar, photographs could be found. Images were sampled from a variety of internet sources (e.g. stock photography, google image search, etc.). Once all these conditions were met, the collection of associated images contained 71 object-context pairs, each with 4 target and 4 context photographs containing unique tokens, for a total of 568 images. Figure 7 shows some of the photographs used as target stimuli and their associated contexts.

Figure 7. Examples of the object-context photograph pairs used in the object recognition task

A comparison of the target and contextual images will reveals several important stimulus properties. The size of the target items are approximately equivalent, ensuring that observers cannot use the size of a briefly presented item to guess its identity. Presenting comparably sized objects also increases the amount of masking by increasing the number of overlapping contours between successively presented stimuli. Secondly, most contextual images are larger than the target items. This ensures that subjects can identify the contextual image as something other than a target. We want our observers to have both the ability and inclination to ignore the scenes to test whether they involuntarily orient towards them.

In the following experiments, if the contextual image on a trial was originally paired with the target object on that trial, it will be referred to as a related associated or related contextual distractor. Similarly, if the contextual

image on a trial was originally paired with a different target object, it will be referred to as an unassociated or unrelated contextual distractor.

Predictions for the Current Experiments

In the present experiments, the hypothesized attentional capture account is as follows. When observers are instructed to monitor a rapidly changing sequence of photographic objects to identify and possibly encode a verbally labeled target object, they do so on the basis of diffuse ACS that includes conceptual knowledge about places where the target object was encountered previously. When they are presented with an image that matches a currently active contextual representation, this distractor competes more successfully with the target than a contextual representation that is not currently active. This line of reasoning leads to the counterintuitive prediction that performance will suffer on trials where associated contexts precede target objects. This prediction is interesting because associated contexts are typically demonstrated to facilitate object recognition (Davenport & Potter, 2004; Auckland, Cave, & Donnelly, 2007). However, as mentioned previously, the facilitative effect of context in these experiments takes place when the target is unknown and must be recollected or identified after viewing.

In Experiment 1, observers were instructed to detect a verbally labelled target object presented within a rapidly presented sequence of images. It has been demonstrated that detection deficits for rapidly presented stimuli depend on the relationship between distractors (here a contextual image) and the attentional

criteria used to identify targets (Leber & Egeth, 2006). In this case, the association between a target object and its preceding context was expected to modulate the likelihood that observers will detect the target object. As predicted, observers were less sensitive to target objects when those target objects followed associated contextual images. Later experiments replicate and extend this cost for discrimination.

Chapter 4: Contextual Capture and Detection

Experiment 1

Observers detected a verbally indicated target object category in a sequence of 12 rapidly presented photographs.

Method

Participants. 34 (22 female) undergraduate students voluntarily participated in this experiment for extra-credit in undergraduate psychology classes. All participants had normal or corrected to normal vision.

Stimuli. Photographs of common objects and scenes, as described in Chapter 3 and listed in Appendix 1, were presented on an LCD monitor in dim light at a distance of approximately 50 cm. Objects varied in size and area. The average size of an object photograph was 11.65° (SD = 4.53) across by 10.61° (SD = 6.12) high. Scene photographs averaged 24.43° (SD = 2.17) horizontally by 19.38° (SD = 3.79) vertically. For each of 69 object categories, there were 4 token images and 4 associated scene images, yielding a total of 552 images used in the experiment. The sampling of these images is addressed in the design section below.

Procedure. Each subject completed 276 self-paced, test trials. As shown in Figure 8, each trial began with the press of the space bar after which

observers were presented with a one or two word verbal label for 500 ms indicating the category of object they were to detect in the image sequence. Immediately following the presentation of the target object category label, a 12 image sequence was presented at 80 ms/item.

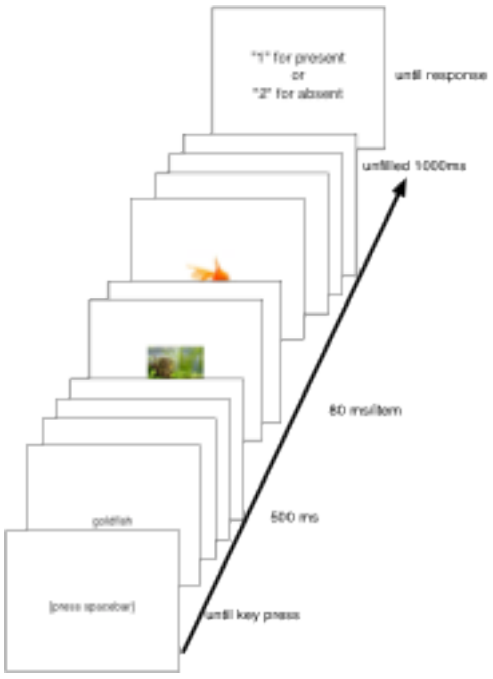

Figure 8. Shows the trial sequence in Experiment 1. Observers indicated whether cued object category was present in the sequence.

After the sequence of images was presented, observers were presented with a blank screen for 1000ms. Following this unfilled period, observers were prompted to indicate whether or not they perceived the target object. If observers believed the target item was presented in the sequence they pressed "1". If they believed the target item was not presented in the sequence they pressed "2". The entire experiment took an average of 20 minutes.

Design. Target-context relationship (associated, unassociated), target presence (absent, present), and target-context relative serial position (lag 1, lag 3, and lag 5) were manipulated. Observers searched for each of 69 object categories four times . On half of trials the verbally cued target was present. Out of those trials where a target was present, it was presented once following an associated context and once following an unassociated context. On each trial, if the target object was present it was represented by a photograph that was chosen randomly without replacement from the 4 possible category token images. Since each target image was used as a target only once, participants could not rely on strategies focusing on local features. Similarly, on all trials contextual images were sampled without replacement from a pool of images. In other words, no contextual image was used in more than one trial across the entire experiment. As mentioned previously, a variable mapping design was employed with the same collection of photographs serving as target and distractors. While each target image was only used as a target once, the entire collection of target images was used as a pool of distractor objects. Images did repeat an average of four times across the experiment as distractors, but only appeared as targets once.

The relative serial position of the context image and the verbally cued target item was set to one of three possible lags. In all of the following experiments, lag refers to the temporal relationship between the sequentially presented stimuli. The lag level of a target present trial describes the relative serial positions of the contextual image and target. For example, lag 1 trials

involve the target object presented immediately following the contextual image, whereas lag 3 trials entailed the presentation of two intervening distractors between the context image and the target. Targets were positioned randomly in serial positions 6 through 10. Contextual images preceded these targets by 1, 3, or 5 positions, meaning contextual images appeared at serial positions 1 through 9. Contextual distractors always preceded target objects. Critically, the preceding contextual image could either be associated or unassociated with the target category. On trials where the target object was associated with the contextual image, it was expected sustained top-down attentional engagement with the distracting contextual image will decrease sensitivity for targets following soon after this distractor. The target-context relationship (associated, unassociated) and target-context lag positions (1, 2, and 3) were crossed within-subjects yielding 6 types of target present trials. It should be noted that on trials without a target present, observers' false alarm and correct rejection responses cannot be associated with any particular lag condition. In other words, there were only two types of target absent trials, those with associated contextual distractors and those without. The overall false alarms following associated or unassociated contextual images will be used along with a particular hit rate to estimate sensitivity at each lag.

Results

Data from one subject was excluded for exceeding low performance (HR = .75, FA = .30, d' = 1.21). Overall performance was high with a HR = .87 (SD = .06) and a FA = .10 (SD = .07). Subjects' sensitivity average d' = 2.78 (SD =

.58). Figure 9 shows hit rates across the lag and contextual distractor relatedness conditions.

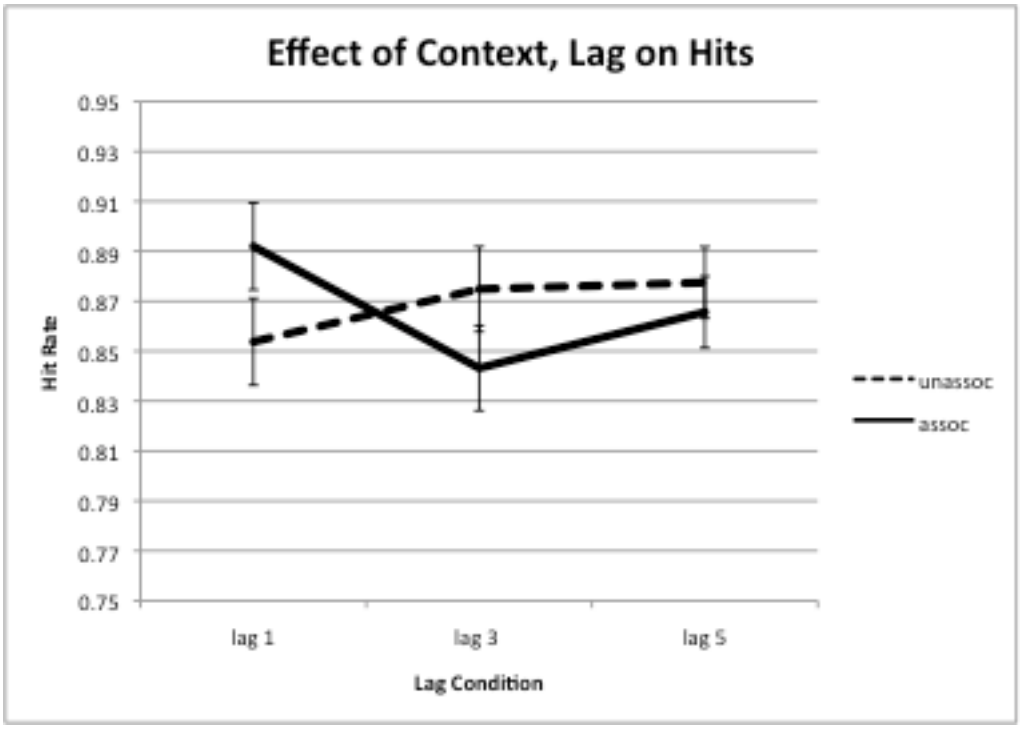

Figure 9. Hit rates for subjects in Experiment 1. Error bars represent the standard error of the context effect at a given lag.

For each subject, sensitivity was calculated at each lag by context crossing. Sensitivity for each of the 6 condition cells was quantified in terms of d' using the hit rate at a given crossing and either the overall associated or overall unassociated false alarm rate. As mentioned previously, trials without targets cannot be associated with any particular context to target lag. Figure 10 shows the data.

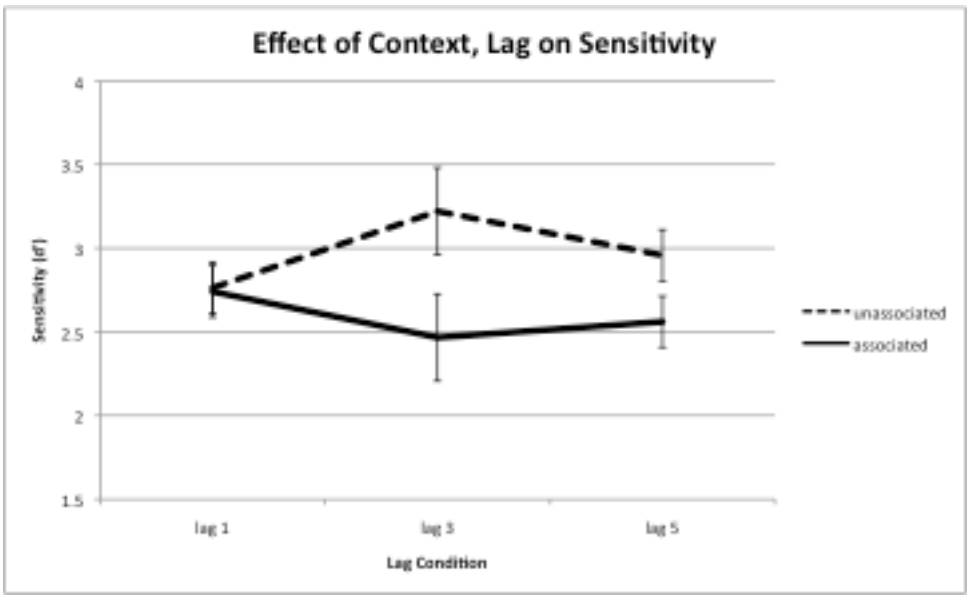

Figure 10. Sensitivity for object targets following associated or unassociated contexts at each lag. Error bars represent the standard error of the contextual effect at a given lag.

Because of the clear relationship between the motivating theory and experimental design, planned repeated measures t-tests were conducted at each lag position comparing sensitivity following an associated or unassociated contextual image without an omnibus ANOVA. Higher values of d' indicate greater sensitivity. At lag 1, observer sensitivity did not differ reliably, $t(32) = .16$, $p > .05$. However, at lag 3 participants performed significantly more poorly on trials containing related contextual distractors ($M = 2.47$, $SD = .18$) compared with trials containing unrelated contextual distractors ($M = 3.22$, $SD = .20$), $t(32) = 2.90$, $p < .01$. Similarly, at lag 5, observers were more sensitive to object photograph targets following unassociated ($M = 2.95$, $SD = .14$) compared with associated ($M = 2.56$, $SD = .14$) contexts, $t(32) = 2.59$, $p = .01$.

Bias was calculated for each subject using either the associated or the unassociated false alarm rate and the hit rate at a given context by lag crossing. Higher values of the criterion estimate C indicate a more conservative standard of evidence for observers. At lag 1, a repeated measures t-test indicates that subjects were significantly more conservative following an unassociated (M = .25, SD = .29) compared with an associated (M = -.07, SD = .55) context, $t(32) = 3.52$, $p < .01$. Differences in bias obtained at no other lag, p's > .05.

Discussion

Overall performance was high; observers were able to quickly detect common objects in the RSVP stream successfully. This indicates that observers understood the task and found it manageable. However, the influence of associated contexts was evident in both sensitivity and bias effects. The costs of the associated context at lags 3 and 5 are likely the best demonstration to date of true top-down attentional capture for the following reasons. First, this demonstration of attentional capture occurs against a backdrop of an unspeeded detection task. Typical experiments in attentional capture use response time as a dependent measure. This is problematic for reasons discussed previously. Secondly, in studies dealing with attentional capture, what is often treated as a top-down effect can be easily explained in terms of inter-trial priming (Maljkovic & Nakayama, 1994; Belopolsky, Schreij, & Theeuwes, 2010). For example, if a subject is instructed to attend to selectively attend to a color and then actually views this color as the focus of attention across some number of trials, how can the effect of intending to attend to red be distinguished from the effect of viewing

red? This experiment can distinguish between these two possibilities. On each trial, observers were required to establish an attentional set different from the previous trial. This likely requires encoding of a verbal label and the generation of anticipatory representations that can be compared with upcoming stimuli. However, this is insufficient to truly establish top-down control of visual attention. If categories or images repeat frequently, manipulations involving trial to trial changes can only measure differences in the magnitude of intertrial priming and not capture in the absence of intertrial priming. In this experiment, observers were presented with a target present trial only twice in the entire experiment. Each of these viewings involved a unique token image from the category. Through this control, the priming influence of one trial with a target category on a later trial with same category is minimized to the extent it is possible while maintaining a fair comparison between conditions. In typical RSVP studies dealing with attentional capture, targets are defined in terms of properties along a single dimension across all trials (Barnard et al., 2004), across blocks (Leber, 2004), across alternating runs (Lien, Ruthruff & Johnson, 2010), or in random sequences (Belopolsky, Schreij, & Theeuwes, 2010). To the best of my knowledge, no study of attentional capture has employed such a broad range of categorically defined targets in an unspeeded perceptual discrimination.

This experiment replicates recent findings showing that associations between objects and contexts results in predictable differences in attentional prioritization (Bar, 2004; Castelhano & Heaven, 2010). More importantly, these data indicate that while observers can establish ACS relevant for a target on a

particular trial, the manner in which an observer's ACS are utilized is partly out of observers' control. In this experiment, there was no advantage for attending to the contextual images in the RSVP sequence on any trial. Contextual images never contained the target item and did not reliably signal the presence of the target item. The difference in size between the contextual and target images was salient and could have been used as a cue to exclude processing of the contextual image. Despite this, scene context images influenced observer performance, such that associated contexts captured attention. Lastly, the costs of attentional capture on sensitivity take at least 80 ms to accumulate, replicating Auckland, Cave & Donnelly's (2007) finding that associated objects sharing a common onset with a target do not affect object recognition performance (but see Joubert et al., 2008). In a related finding, object-based spatial attention capture effects are greatest when the object precedes the appearance of the cue and the target (Shomstein & Behrman, 2008). Reliable differences in sensitivity obtained only at lags 3 and 5.

It is difficult to see how the effect of the contextual distractor could be attributed to any low-level sensory difference between the associated and unassociated contextual scenes. In fact, the design ensured that the same images that appeared before related targets for one subject appeared before unrelated targets for others. In terms of the overall perceptual difference between the contextual image and succeeding targets, one can see a masking effect for targets following either associated or unassociated contexts at lag-1. A large low-level perceptual contrast in an RSVP sequence has been

demonstrated to disrupt attentional performance. For example, recent evidence from Asplund et al. (2010) indicates that observers will miss targets in RSVP sequences due to an orienting response towards novel distractors. When observers were presented with an unexpected stimulus, a deficit similar to the AB obtained for targets following soon after this unexpected distractor. While the contextual scenes were novel on each trial and dissimilar from other object images in both scale and complexity, it is unlikely that the capture effect observed here is due to the surprise induced blindness described by Asplund and colleagues. The surprise effect described in their study persisted only through the first few times that observers encountered an unexpected stimulus. By the time the observers reached their third surprise trial, the capture costs reversed. In this experiment, a single novel contextual scene was shown on every trial, making it unlikely that an orienting response would persist over the course of the experiment. More importantly, differences here are between related and unrelated contexts of equivalent novelty because each image was only viewed once. Any differences in performance must be due solely to the relatedness of the contextual image.

This experiment is wholly consistent with the hypothesis that when observers are searching for an object they maintain anticipatory representations of schematic contexts associated with the target object. This entirely top-down attentional capture effect left a 300 - 500 ms interval within which subjects were less sensitive to images of the target object. However, interpretation of this effect is complicated by the use of detection as a dependent measure. Strategic

guessing biased by scene information is easily mistakable for facilitation of schema consistent items (Hollingworth & Henderson, 1998; Auckland, Cave, & Donnelly, 2007). Analytic techniques such as signal detection theory can correct for bias mathematically, but careful experimental design can yield observations where the influence of observer bias is minimized (Pelli & Farell, 1995). This is particularly important given the debated role of bias in understanding the influence of context on object recognition (Biederman, 1981; Hollingworth & Henderson, 1998).

As mentioned in the discussion of Moores and colleagues work (2003), detection measures give little information about what types of information observers are using to complete the task. In order to more precisely characterize this attentional capture effect, observers will need to perform a discrimination between two simultaneously presented targets. The following chapter describes a series of experiments using discrimination accuracy as a performance measure.

Chapter 5: Contextual Capture and Discrimination

In the following experiments, rather than having observers report whether or not a target item was presented in the sequence, observers indicated which of two category tokens was present in a two alternative forced choice (2AFC) task. Because both of these items fit the verbal description of the target item provided at the start of trial equally well, bias effects that might arise from guessing on the basis of the associated or unassociated contextual image were minimized. Any differences in performance are unlikely to be due to observers' response strategies because observers issued a single, unspeeded response discriminating between two alternatives that were equivalent along independent variable levels (Grider & Malmberg, 2008; Zeelenberg, Wagenmakers, Rotteveel, 2006). On average, observers should have no prior reason to choose one target token over another in the presented comparisons. For example, if observers are instructed to find a chair in a sequence of images and two chairs are presented at the end of the trial, one present in the sequence and one a lure, there is no reason why either the verbal label displayed at the beginning of the trial or the contextual scene should bias observers towards one chair or another.

While presenting two alternatives does eliminate criterion setting bias, there are additional subtle differences between detection and discrimination tasks. These differences make the contextual distractors less likely to have an

effect, so any positive evidence of contextual costs in these experiments can be interpreted as strong evidence in support of the motivating hypothesis. First, rather than simply compare each presented image with anticipatory representations and retrospectively respond at the end of a trial when a match is detected, observers must now evaluate each item for relevance, make an online decision about category status, and selectively encode the visual details of matching items. Previous research on attentional and perceptual load suggests that distractors are less likely to be processed when target processing is complex (Lavie, 1995). To the extent that a within category discrimination is more perceptually complex than a simple detection task, it is less likely that contextual distractors will be engage perceptual processes. Apart from these general task concerns, the distinguishing details for within category discriminations are concentrated at high spatial frequencies. ACS in rapid picture perception include selection on the basis of spatial frequency information (Schyns & Oliva, 1997). Schematic scene categorization is closely associated with information concentrated at low spatial frequencies (Oliva & Torralba, 2001). If subjects are selectively attending to information concentrated at high spatial frequencies, this may reduce the influence of the schematic contextual scene. Lastly, discrimination is often an easier task than detection because the observer is presented with the target a second time. Given these three considerations that might mitigate capture effects, if the influence of associated contexts is observed in Experiment 2, this would be stronger evidence that the processing of associated contexts is forced by diffuse ACS.

Experiment 2

Method

Participants. 43 (29 female) undergraduate students voluntarily participated in this experiment for extra-credit in undergraduate psychology classes.

Stimuli. The stimuli from Experiment 1 were used. Viewing conditions were identical to Experiment 1.

Procedure. Each subject completed 138 test trials viewing each of the 69 object context pairs twice. The experiment was self paced. As shown in Figure 11, each trial began with the press of the space bar after which observers were presented with a one or two word verbal label indicating an object category they were to selectively encode from the image sequence for 500 ms. Immediately following the presentation of the target object category label, observers were presented with a 12 image sequence at 100 ms/item.

After the sequence of images was presented, observers viewed a blank screen for 1000 ms, followed by two tokens from the target category. Observers then indicated which of the simultaneously presented tokens was present in the RSVP sequence. Since both objects were drawn from the same verbally labeled target category, observers had no prior reason to select either of the choices. Responses were captured using spatially mapped buttons, such that if observers believed the left item was presented in the sequence they pressed “1” and if they believed the right item was presented they pressed “2”. Targets and lures were

presented randomly as either the choice on the left or the choice on the right. The entire experiment took an average of 15 minutes.

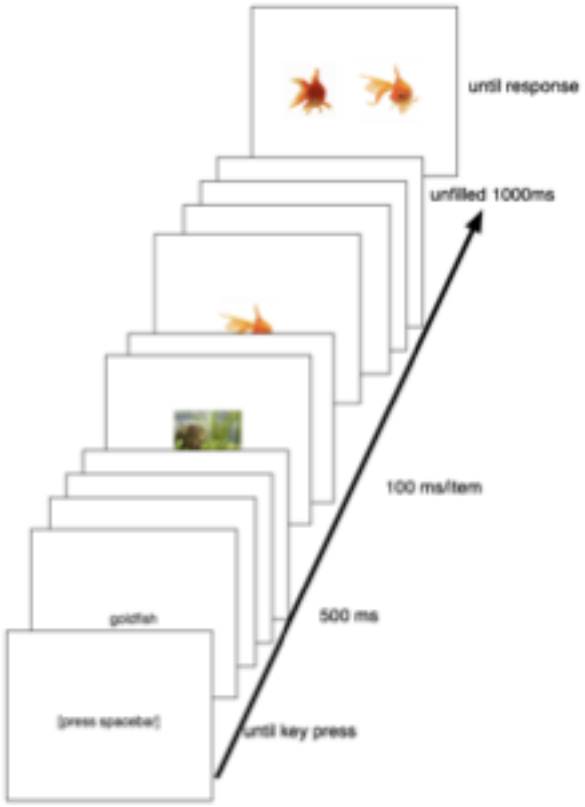

Figure 11. Shows the trial sequence in Experiment 2. Observers indicated which of two tokens was presented in a trial.

Design. Contextual distractor relationship (associated, unassociated) and lag (lag 1, lag 2, lag 3) were manipulated. These sequences contained one instance of the target object category which was inserted randomly into serial positions 5-8. This target object was preceded by a context image at either 1, 2, or 3 serial positions prior (serial positions 2-7). As in Experiment 1, this distracting contextual image was either associated or unassociated with the

target object. It is anticipated that observers will be less accurate when the verbally cued target follows soon after an associated context.

On each exposure to any object context pair, participants were presented with a novel target object or context token that was chosen randomly without replacement. No picture used as either a target or a lure in a 2AFC discrimination more than once. However, as with the previous experiment, the same collection of images was used as both targets and distractors to control for image novelty and statistics.

Results

Observers correctly identified the presented target category token on 79.1% (SD = 5.4%) of trials. As can be seen in Figure 12, performance varied across both lag and contextual image relatedness conditions. Three planned comparisons were conducted with t-tests for each lag condition comparing performance with related or unrelated contextual images. At lag 1, no reliable differences obtained when comparing target recognition accuracy following related (M = 78.9%, SD = 9.5%) and unrelated (M = 79.7%, SD = 10.1%) contextual images, $t(42) = .38$, $p > .05$. A second planned t-test compared related (M = 76.9%, SD = 10.5%) and unrelated (M = 81.3%, SD = 9.2%) performance at lag 2, revealing a reliable cost for items following closely after an associated context, $t(42) = 2.20$, $p = .03$. Similarly, a planned t-test at lag 3 revealed costs for associated contexts (M = 77.3%, SD = 7.9%) over unassociated ones (M = 80.6%, SD = 8.3%), $t(42) = 2.18$, $p = .03$.

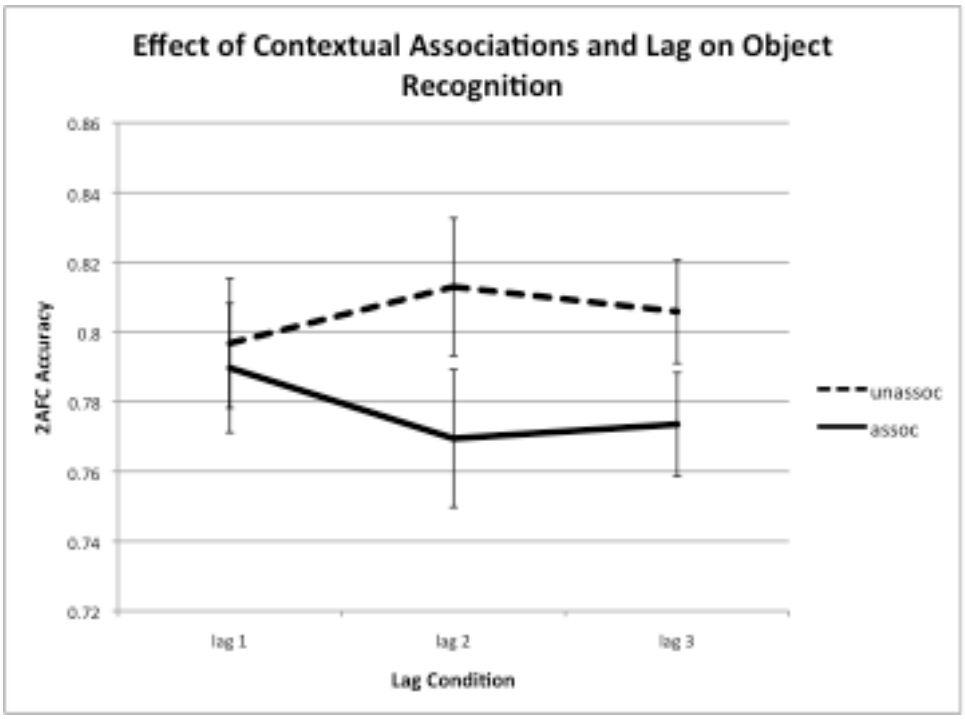

Figure 12. 2AFC accuracy as a function of the preceding contextual image and the lag condition. Error bars represent the standard error of the contextual effect at a given lag.

Discussion

As in Experiment 1, observers were able to rapidly establish anticipatory representations that permitted the selective treatment of target items in a stream of rapidly presented images. In fact, a cursory comparison of Figures 13 and 11 reveals a very similar effect. In both cases, between costs from preceding contexts do not obtain unless targets fall at serial positions greater than lag 1. Unlike Experiment 1, here observers were instructed to make subtle within-category discriminations comparing two similar tokens. Observers correctly indicated the presented token on nearly four fifths of all trials. This reflects

considerable flexibility in visual ACS representations. These anticipatory representations can not only be used to detect target items in a rapid sequence, but can also be used to trigger the selective encoding of distinguishing token features. Further, given that subtle differences in pose, color, texture, and form distinguished between the target and lure on each trial, these data are evidence that relatively detailed object representations are constructed by observers. Moreover, the costs indicate that the encoding of this detailed object knowledge is disrupted these associated contexts.

Despite this overall high discrimination performance, observers showed costs when targets were preceded by associated contexts compared with unrelated contexts. Moreover, as with detection, this effect only emerged at lag positions greater than 1. This replicated pattern of effects shows that associated contexts engage perceptual processing when observers search for common objects in a way that unassociated contexts do not and that this capture takes time to harm recognition performance. Not only does this finding extend the generality of the effect from Experiment 1, it also demonstrates the effect in a task domain where observers were performing a discrimination relying on high spatial frequency information. As mentioned previously, observers can tune ACS to selectively encode spatial frequency ranges appropriate for a task (Schyns & Oliva, 1997). Information that is diagnostic for scene category is generally thought to be concentrated at low spatial frequencies (Bar, 2004). The fact that relatively low spatial frequency information influenced the selective encoding of

high spatial frequency information is suggestive of obligatory processing of related scenes.

To describe the effect of related contexts more specifically, observers were less likely to accurately identify a target object when that target object followed associated contexts at lags greater than 1. As mentioned previously, lag 1 sparing is a diagnostic feature in attentional blink phenomena. The importance of lag 1 sparing, the phenomena's underlying mechanisms and boundary conditions are currently debated (Dux & Marois, 2009). The meaning of lag 1 sparing in these particular experiments is uncertain because the meaning of the effect is only coarsely characterized in general.

However, the interpretation of Experiments 1 and 2 is complicated by the fact that a scene, either related or unrelated to the target, was presented on each trial. This was a key manipulation in demonstrating that associated contexts capture attention. If only a portion of trials contain contextual distractors, these distractors will have increased salience due to their greater novelty. This might result in a bottom-up capture effect. In Experiments 1 and 2, we see costs of the distractors despite the fact that they are present on every trial. At the same time, this means that the effect of the contextual manipulation cannot be attributed to either the enhancement of target processing following unrelated contexts or the impairment of target processing following related contexts. Previous research has shown selection of items that are inconsistent with a simultaneously presented scene context (Hollingworth & Henderson, 2000; Gordon, 2006). The difference in the effect of the contextual distractor as we move from lag 1 to lag 3

could be the enhanced processing of targets following a mismatched context. Alternatively, this difference could be the release of a target from forward masking following an perceptually dissimilar unassociated context image. That is, both the target following an associated context and the target following an unassociated context are subject to forward masking at lag 1. At lags greater than 1, the meaning of the stimulus is processed in the case of the target following an associated context, leading to reduced performance. There is no such sustained engagement in the other case. To test the attentional capture hypothesis, in Experiment 3 there are trials in which a contextual distractor is not present. This permits the comparison of performance for targets following no contextual distractors, related contextual distractors, and unrelated contextual distractors. If an unrelated context affords an encoding advantage for a target on a trial containing a contextual distractor compared with a trial without any contextual distractor, then the effect observed in Experiments 1 and 2 is due to the mismatch between the contextual distractor and the succeeding target. On the other hand, if the attentional capture hypothesis is correct, we would only expect to see a difference between the related context trials and the no context trials.

There are other issues in the first two experiments that need to be addressed. In order to describe an effect as attentional capture, it is necessary to describe performance of the task both before and after costs are observed. While additional chronometric exploration is desirable to fully characterize capture effects, at a minimum one must show performance before capture,

during capture, and following capture. By measuring performance after a return to baseline, one ensures that the effect is temporally circumscribed and not the result of durable changes in some other cognitive mechanism (e.g. decision-making). In Experiment 3, lag 5 trials are replaced with lag 6 trials in an effort to demonstrate recovery following capture. Lastly, while we have asserted that an unspeeded discrimination should not result in speed accuracy trade-offs, we have not measured response time. In Experiment 3, we will measure response time directly to determine whether observers might respond more quickly in a given a condition, resulting in poorer performance than would otherwise be the case.

Lastly, in Experiments 1 and 2 great pains were taken to prevent the repetition of target images. This decision, along with the variable mapping design, was made to minimize the possible influence of intertrial priming or the selective encoding of diagnostic features. However, to show capture even when specific target images repeat extends the generality of the effect to cases where these possible influences are present. In Experiment 3, target images were randomly sampled with replacement.

Experiment 3

Method

Participants. 25 (23 Female) undergraduate observers with normal or corrected vision participated in the experiment. Observers were given extra-credit for participation.

Stimuli. The same stimuli from Experiment 1 and 2 were used a third time. Viewing conditions were identical to Experiments 1 and 2.

Procedure and design. Observers completed 414 test trials identifying tokens from each of 69 object categories in a within-subjects factorial design. Each target category appeared 6 times during test trials, with two trials following no context, two trials following an unassociated context, and two trials following an associated context. Both target images and lures were chosen randomly with replacement. Before participants began these test trials, they completed 24 practice trials with no contextual distractors using the same set of images as subsequent test trials. Including both test and practice trials, each image was used approximately 18 times as a distractor in an RSVP sequence.

The timing of a given trial in this self-paced experiment was identical to Experiment 2. As mentioned, the context-target interval now contained lags 1, 3, and 6. Also, on a third of trials, the target object was presented without a distracting contextual image. This establishes a baseline performance for comparison to the experimental conditions. As before, the relationship between the preceding contextual image, on the two thirds of trials that contained contextual distractors, and target image was manipulated so that it could either be associated or unassociated. Observers completed the same bias-controlling 2AFC at the end of each trial as in Experiment 2. The entire experiment took an average of 45 minutes for participants to complete.

Results

Figure 14 shows overall object recognition performance on trials containing contextual distractors in Experiment 3. Because lag conditions are not meaningful when contextual distractors are absent from a trial, the effect of context will be measured while collapsing across lag conditions. A one-way within-subjects ANOVA with context condition (none, unassociated, associated) as the independent variable revealed a reliable effect of context, $F(2,48) = 3.414$, $p = .04$. A planned within-subjects t-test comparing control ($M = 84.7\%$, $SD = 6.9\%$) and unrelated ($M = 83.1\%$, $SD = .05\%$) conditions revealed no reliable effect of unrelated contexts, $t(24) = 1.46$, $p = .16$. On the other hand, a planned within-subjects t-test comparing control and related ($M = 82.1\%$, $SD = 5.4\%$) conditions, showed an advantage for targets on trials without contextual distractors, $t(24) = 2.39$, $p = .03$. These data indicate that the effects observed in Experiments 1 and 2 are costs of related contexts, rather than an advantage for items following unrelated contexts.

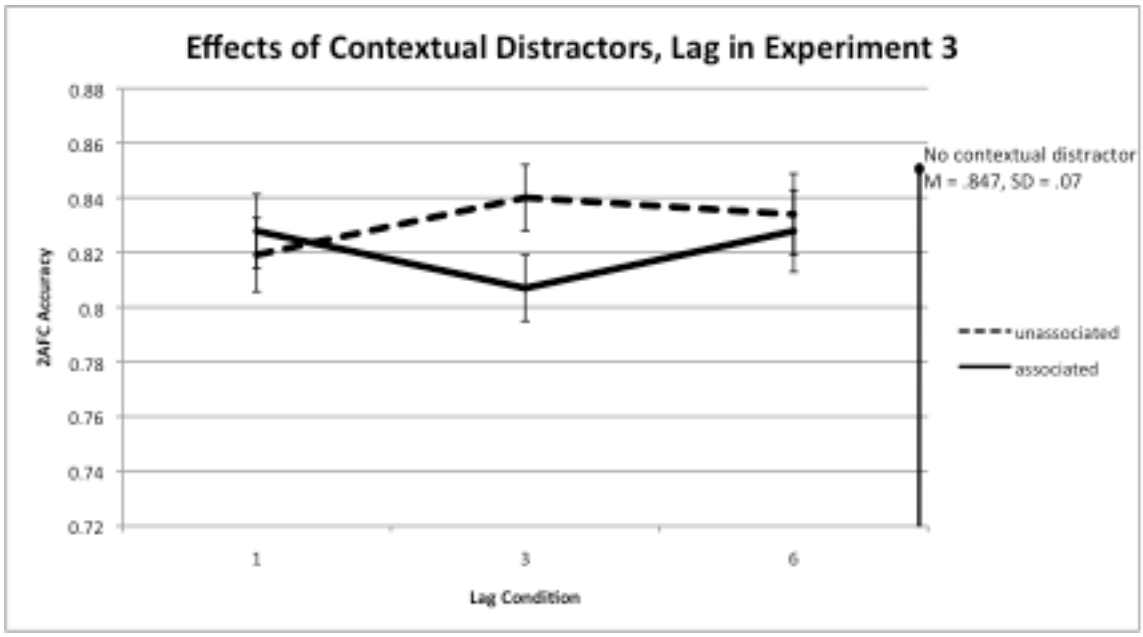

Figure 13. The effects of contextual distractors and lag in Experiment 3. Accuracy on contextual distractor free, and hence not lag conditioned, trials is visualized with the plot on the right edge. Error bars indicate the standard error of the contextual effect at a given lag.

Planned within-subjects t-tests were conducted at each lag condition. As was the case in Experiments 1 and 2, there was no difference in performance at lag 1, $t(24) = .61$, $p = .54$. In a replication of Experiment 2, at lag 2 there was a cost for targets following related contexts ($M = 80.7\%$, $SD = 6.9\%$) compared with unrelated contexts ($M = 84.0\%$, $SD = 6.3\%$), $t(24) = 2.82$, $p = .01$. Last, and importantly, there was no reliable difference between discrimination accuracy for targets following associated and unassociated contexts at lag 6, $t(24) = .40$, $p = .70$. This recovery for targets occurring late in the sequence demonstrates that the cost following the related context is transient. This is consistent with an attentional capture account.

As mentioned previously, response time measures were captured to ensure that observers were not engaging in a speed-accuracy trade off. A one-way, within-subjects ANOVA treating context (none, unassociated, associated) revealed no reliable effect of context on response time, $F(2,48) = .07$, $p = .94$. That is, response times following no context ($M = 1074$ ms, $SD = 25$ ms), an unrelated context ($M = 1069$ ms, $SD = 29$ ms), and a related context ($M = 1078$, $SD = 31$ ms) were not statistically distinguishable.

Discussion

These data replicate the general pattern of performance observed in Experiments 1 and 2. Namely, costs for targets following distractors did not obtain at lag 1 while these effects were present at lags greater than 1. However, unlike Experiments 1 and 2, target images were repeated in the current experiment. These capture effects do not depend on unfamiliarity with a particular image. We can see this capture effect persists under a broad range of perceptual tasks including detection of named targets, discrimination of relatively unfamiliar targets, and the discrimination of familiar targets.

Experiment 3 did not simply replicate and extend Experiments 1 and 2, but addressed possible deficiencies in their designs. First, it is now reasonable to assert that the effects in Experiments 1 and 2 are costs for targets following related contexts. In the present experiment, there was no reliable difference between overall performance on trials with unassociated contexts and trials without contextual images. In contrast, there were reliable differences between

trials without contextual images and trials with related contexts. This is consistent with an account where related contexts capture visual attention and deprive targets in a circumscribed period following the context of encoding resources. Second, Experiment 3 showed recovery following the contextual distractor at lag 6. The sparing of targets at lag 6 indicates that the related contexts have a transient effect and do not disrupt cognitive mechanism associated with decisions or responses. Third, response time measures from Experiment 3 indicate that observers take approximately the same amount of time to respond on trials without contextual images, those with related contextual images, and those with unrelated contextual images. In fact, while the difference was not statistically significant, subjects were the slowest following related contexts. Since this is the condition in which performance was worst, subjects likely did not trade accuracy for speed.

However, interpretation of the data from Experiments 1, 2, and 3 is complicated by the fact that multiple lags were employed. When observers are presented with a temporal selection task with multiple possible target lags, it can be difficult to adopt a strategy for responding that covers all possible lags equally well (Caetta & Gorea, 2010). For example, an observer might complete 3 trials with a long lag between context images and targets. Over these three trials observers may calibrate their decision processes for a certain level of evidence or adopt a certain encoding strategy. On the fourth trial observers might be asked to identify a second target after a short lag. Poor performance on this fourth trial would then result from two causes. One is the innate change in

difficulty when multiple task relevant items are presented closer in time. This is of primary theoretical interest, will likely make the task easier or harder, and is the intended effect of the manipulation. In addition to this essential cause, strategy carryover effects may also interfere with observer accuracy. Because multiple lags were employed in Experiments 1, 2 and 3, poor performance might have been exacerbated by inappropriate criteria or strategy on particular trials. Another way of thinking about the effect in the first three experiments, involves focusing on the factorial combination of context and lag conditions. If a participant is trying to learn how to report the target and exclude the contextual scene from processing, this will be harder when the temporal relationship between these items is varied. In some strained sense, we are asking the participant to learn a unique task at each of the possible lags. While it is unlikely, the failures of attentional control that results in costs on associated trials could result from control mechanisms being overwhelmed by the frequent "task-switches" as we move from lag to lag. A stronger test of the hypothesis that associated contexts capture attention during temporal search would involve the presentation of contextual images and target images in a fixed temporal pattern. In this way, participants will be able to settle into a consistent internal strategy for identifying targets. In Experiment 4 observers were presented with targets following contexts at a fixed temporal interval

Experiment 4

Method

Participants. 20 (13 Female) undergraduate observers with normal or corrected vision participated in the experiment. Observers were given extra-credit for participation.

Stimuli, design, and procedure. The stimuli from Experiment 1, 2, and 3 were used a third time. Viewing conditions were identical to the previous experiments. These 69 stimuli were supplemented with two more, for a total of 71 object categories . Observers completed 142 trials identifying tokens from each of 71 object categories twice. No image served as a target or lure more than once.

This self-paced experiment was identical to Experiment 2, except that only lag 2 was utilized. As before, the relationship between the preceding contextual image and target image was manipulated so that it could either be associated or unassociated. Observers completed the 2AFC discrimination at the end of each trial as in Experiment 2. The entire experiment took an average of 15 minutes.

Results

Figure 14 shows object recognition performance in Experiment 4. As anticipated, observers were more accurate in identifying targets that appeared following unrelated contexts (M = 77.6%, SD = 5.5%) compared with related contexts (M = 74.1%, SD = 7.2%), $t(19) = 2.43$, $p = .03$. To ensure that

observers' strategy did not change over the course of the experiment, a comparison of attentional capture effects (unrelated context accuracy - related context accuracy) in the first (M = 3.6%, SD = 10.7%) and second halves (M = 4.0%, SD = 12.0%) of the experiment was conducted and revealed no reliable differences, $t(19) = .09$, $p > .05$.

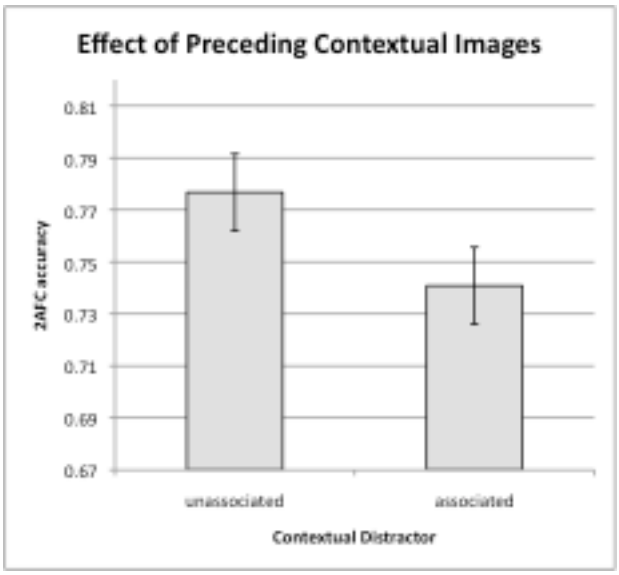

Figure 14. 2AFC accuracy for object photographs following associated and unassociated contexts at lag 2 exclusively. Error bars represent the standard error of the contextual effect.

Discussion

Experiment 4 demonstrates that even in situations where observers can set a uniform encoding strategy with a highly practiced perceptual task, costs still obtain for target items following associated contexts. This attentional capture effect is so robust that even when a single lag is used across the entire experiment observers show no ability to overcome this cost. Additionally, the

capture effect was uniform between experimental halves. This argues against explanations based around participant misunderstanding or speech pragmatics. If observers mistakenly believe they are to find the target object located in a contextual scene, they should be disabused of this notion by the time they complete the 71st trial. Despite what is likely to be a clear understanding of the task, observers can't help but attend to the contextual images.

Experiment 5

There are at least two explanations that might account for the contextual costs demonstrated in the previous four experiments. On the one hand, I have been advocating an explanation based on attentional capture. From this perspective, observers' performance suffers because they are identifying relevant information in the RSVP stream on the basis of diffuse ACS that includes representations of contexts where target items have been previously encountered. When the presented images match this context, perceptual processing is engaged by the contextual image and performance suffers because, by the time the target is presented, insufficient resources are available for elaborated representation. Alternatively, these effects might also be explained in terms of interference between related items in near term memory processes. That is, we could be observing something analogous to a failure in directed forgetting. In an account focused on interference effects, related contexts may simply compete with targets more effectively than unrelated contexts. Observers' poor performance in the related condition doesn't necessarily result from capture directly, but all distractors are processed to the

degree they match anticipatory representations. On trials with related contexts this processing continues further than on trials with unrelated contexts. While both of these possibilities involve breakdowns in attentional control on the basis of object-context associations, there are important differences. If interference between related items in visual short-term memory (VSTM) or conceptual short-term memory (CSTM) explains the effect we would anticipate similar costs for targets that are followed by associated contexts in much the same way we see costs for targets that precede contexts. However, if attentional capture has a role in this related context cost, it would be important that the context precede the target in the RSVP sequence.

Method

Participants. 34 (26 Female) undergraduate observers with normal or corrected vision participated in the experiment. Observers were given extra-credit for participation.

Stimuli, design, and procedure. The experimental design was exactly the same as Experiment 4, with associated and unassociated object-context pairs presented in rapid succession in RSVP sequences. The only difference between Experiment 4 and Experiment 5 involves the order in which the associated items were presented. In Experiment 4, the context preceded the object, in Experiment 5 the object will precede the context. Because of this manipulation, objects appeared on average two serial positions earlier in Experiment 5 than in Experiment 4. Previous research has identified a cost for

items occurring early in an RSVP sequence, referred to as an attentional awakening effect (Ariga & Yokosawa, 2008). However, the difference in target serial position between Experiments 4 and 5 is small. Further, to the extent that this experiment is designed to provide a strong evaluation of the attentional capture account, if subjects perform poorly in Experiment 5 due to serial position differences that is not a difficulty because we would predict better performance. More importantly, our comparison addresses the associative relationship between the target and context, so differences in overall performance are not important provided they do not push participants to a performance floor or ceiling.

Results

Discrimination performance in shown in Figure 15. As anticipated a within-subjects t-test did not reveal any differences in discrimination accuracy for targets following unassociated (M = 79.1%, SD = 4.7%) or associated (M = 80.4%, SD = 6.9%) contexts, $t(33) = 1.21$, $p = .23$. While null results must always be interpreted with caution, a post-hoc power analysis assuming the effect size of Experiment 4 indicates that observed power was approximately .93. This is a reasonable level of power. An independent samples t-test comparing overall accuracy in Experiment 4 (M = 75.8%, SD = 5.4%) and Experiment 5 (M = 79.8%, SD = 4.9%), $t(53) = 2.70$, $p = .01$.

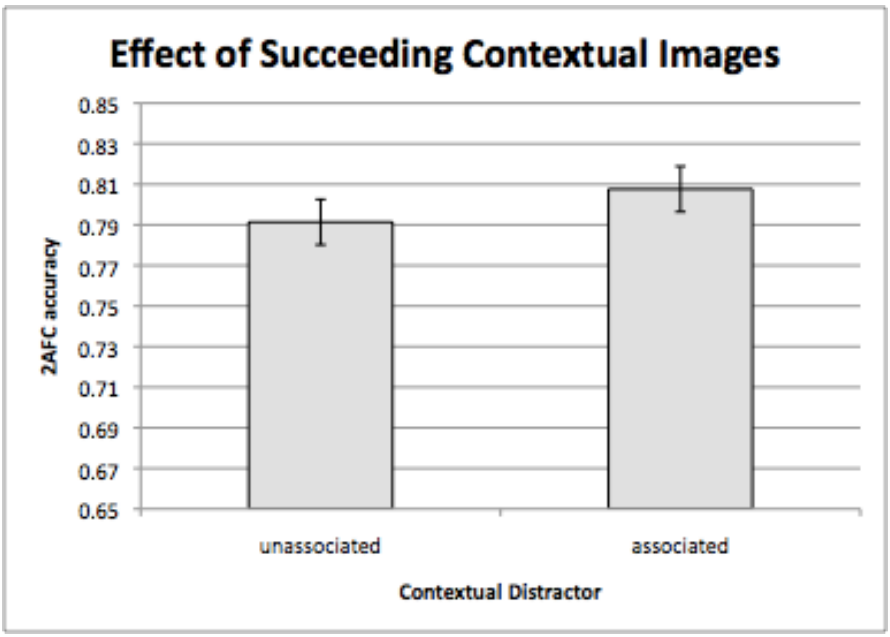

Figure 15. 2AFC accuracy for object photographs followed by associated and unassociated contexts at lag 2. Error bars represent the standard error of the contextual effect.

Discussion

Experiment 5 provided no evidence of costs for targets that are followed by related contexts. Moreover, overall performance was higher in Experiment 5 than Experiment 4. These two findings are consistent with an account based on attentional capture and not interference in memory during the retention interval between the presentation of the target and the discrimination response at the end of a trial.

Chapter 6: General Discussion

These experiments were motivated by the hypothesis that depictions of schematized associated contexts capture visual attention during object search. Specifically, it was hypothesized that this attentional capture effect is non-spatial, transient, and will result in costs for targets following soon after a related context. While attending to contexts associated with a target is typically adaptive, in these experiments the presentation of the target of the target and context was manipulated to set the two in competition in a challenging object search task. Targets compete effectively with associates (Moores et al., 2003), so structuring a task where costs are observed presents challenges. These experiments used a focal contingent capture paradigm (Ghorashi et al., 2003) where both distractors and targets appear in rapid succession at fixation.

The following discussion will review the implications of these experiments for our understanding of the control of visual attention and object-context relationships in object identification. First, we will review the key findings of these experiments. Second, we will review the consequences of these experiments for theories of the control of visual attention. Third, we will address the role of contextual associations in object recognition, as illuminated by this line of research.

Key Findings

Experiment 1 demonstrated that in a detection task, observers' sensitivity is harmed when the target is preceded by an associated context. In Experiment 2, we replicated this cost using a discrimination task. Detection and discrimination are related, but distinct perceptual tasks (de la Rosa, Choudhery, Chatziastros, 2011). Experiment 3 addressed possible methodological concerns from Experiments 1 and 2. Specifically, we demonstrated that the effect of the contextual manipulation is a cost for items following soon after associated contexts (rather than a benefit for targets following unassociated contexts). Additionally, this experiment indicates that this cost is transient, as recovery was observed for targets appearing in the lag 6 condition. Lastly, this experiment indicates that even when observers are repeatedly shown the same target images, performance for targets following associated contexts still suffers. Experiment 4 indicates that even when the targets and contexts are presented in a stable temporal relationship, costs from associated contexts still obtain. Finally, in Experiment 5, we show that these costs are likely due to encoding processes. When observers are presented with the same items in the opposite order, that is, when the contextual distractor follows the target, there was no cost of an associated context.

These five experiments describe the role of abstract conceptual knowledge about scenes associated with common objects in a temporal search

task. The data clearly demonstrate that: a) knowledge about contextual associations is active during an object search task where this knowledge is irrelevant; b) observers are able to select particular objects as the focus of selective mechanisms, but have limited control over the way in which they search for these objects; c) the effect of presenting a scene that matches this contextual knowledge depends critically on the timing between the presentation of the scene and the target; d) these costs obtain in both detection and discrimination tasks; and e) these capture effects are non-spatial. Methodological controls ensures that each of these conclusions does not depend on strategic guessing or post-perceptual cognitive processes.

Implications for Attentional Control Processes

These experiments are among the best demonstrations to date of top-down attentional capture for several reasons. First, participants completed an unspeeded object search task. While demonstrations that show capture using response time measures are useful (e.g. Moores et al., 2003), one cannot be certain that only perceptual processes are being strained. It is preferable, when evaluating theories of attentional control, that measures of attentional performance load onto data-limited and not resource-limited processes (Norman & Bobrow, 1975). Second, in these experiments the target changed on each trial and repeated infrequently. Recent reports have disputed the role of top-down attentional control in contingent capture paradigms (Belopolsky et al., 2010). In most previous demonstrations using a similar design, the target defining criteria has been blocked across trials. This design choice results in capture effects that

might result from either top-down attentional control or bottom-up priming of task relevant features (Folk & Remington, 2008). In these experiments, because the target category changed on each trial (Experiments 1, 2, 3, 4, and 5) and repeated only once using novel pictures (Experiments 2 and 4), intertrial priming is an unlikely explanation for the capture effects observed. The possibility of intertrial priming was further reduced by the use of a variable mapping design and novel contextual distractors (Experiments 1,2, and 4). Third, attentional capture was demonstrated in paradigms with distinct perceptual tasks. Costs for targets following associated contexts obtained in both detection and discrimination tasks showing the generality of the effect. Fourth, performance for targets suffered in experiments where contextual distractors (both related an unrelated) were present on all trials (Experiment 2) and when the contextual distractors appeared on only two thirds of trials (Experiment 3). This suggests that attentional capture does not depend on the novelty of the contextual distractor in the object RSVP sequence.

These data support accounts where attentional capture can occur on the basis of high-level task representations. Recent demonstrations of attentional capture in visual search, suggests that relational properties (e.g. redder), rather than individuated dimension levels (e.g. red) support selective processing (Becker, 2010). Observers were able to effectively establish search templates that identified categorically defined targets rapidly. However, the specificity of these search templates is low, such that scenes associated with the target

engage processing. Despite extensive practice, observers seem unable to tune their ACS more narrowly and exclude these related distractors.

As mentioned previously, this stands in contrast to recent demonstrations of surprise induced blindness by Asplund and colleagues (2010). In these experiments a novel and visually salient distractor produced deficits in target processing over the course of several hundred milliseconds following its presentation. However, with repeated exposure to these unexpected distractors, participants were able to tune their attentional control mechanisms to exclude this salient, but irrelevant, information. In Experiments 1, 2, and 4, observers were presented with related scene distractors on each of over one hundred trials and these costs persisted.

There are many studies that have shown that participants will attend to emotional stimuli (Codispoti, Bradley, & Lang, 2001). In some ways, the demonstrations in the current experiments are similar to the attentional capture findings described by Most, Chun, Widders, and Zald (2005). In this series of experiments, observers were instructed to locate an oriented landscape in an RSVP sequence containing a variety of upright landscape photos. At the end of the trial, they were to indicate whether the oriented landscape faced the left or the right. Critically, this landscape photo was preceded by an emotionally engaging scene by either 2 or 8 serial positions. When an engaging scene preceded the target by 2 serial positions, discrimination accuracy suffered. There were no costs when this scene preceded the target by 8 serial positions. This deficit is described by the authors as emotion-induced blindness. The

researchers explore a number of variables that influence this effect, including the specificity of the attentional set for the target (any oriented image vs. a particular oriented image) and personality variables (harm avoidance). These experiments demonstrate that when an item engages attention, performance for items following soon after suffers. However, unlike the demonstration by Most and colleagues, in the experiments presented in this paper, observers do not demonstrate a deficit for an overall and consistent class of stimuli (e.g. emotional scenes). Rather, it was the specific relationship between the target and preceding distractor on a particular trial that determined performance. If one takes the perspective that emotion-induced blindness is caused by attentional allocation on the basis of visual cues to situations of biological relevance, it would be fair to characterize attention to emotional stimuli as a persistent and ubiquitous set of ACS parameters. In contrast, the current experiments measures the costs associated with incorrectly allocating attention to scenes on the basis of transient ACS parameters. Whereas Most and colleagues show evidence of emotion-induced blindness, these data might be characterized as task-induced blindness. Task here is defined quite narrowly, as the search for a particular categorically defined target. Participants only miss the targets when they are preceded by a distractor related to that target.

This deficit for targets following related distractors is not without precedent. As mentioned previously, work by Barnard and colleagues (2004) showed that when observers are searching an RSVP sequence for profession words (e.g. baker), targets appearing soon after words that describe non-

professional roles (e.g. father) suffer. Observers apparently maintain a coarse attentional set and words that partially match the target category are selected for elaboration. However, in this previous work observers detected or identified a single target category over the course of the entire experiments. For reasons described earlier, it is easier to demonstrate attentional capture when a single target category is employed for multiple, successive trials. In contrast, the experiments presented in this paper involved a changed target category on each trial. The fact that attentional capture on the basis of related meaning has been demonstrated with both constant and changing target categories suggests that the coarseness of the attentional filter does not depend critically on target category variability.

A related demonstration is provided by Evans & Wolfe (2010). In this series of experiments, participants were instructed to detect a cued scene category in a rapidly presented series of patterned masks. The verbal label used to cue a category was provided either before or after the sequence of images. A limited range of scene categories were employed. When observers were presented with the cued target scene within 200 ms of a meaningful scene from one of the other categories, performance for the target suffered. For example, if observers were instructed to detect a beach scene in the sequence, and this beach scene was preceded by a bridge scene among otherwise meaningless patterned images, subjects were less likely to detect the beach scene because of interference from this other meaningful scene category. It appears as though observers are unable to shield their processing of the cued target from the

interference created by this other meaningful scene, in much the same way that target processing was disrupted in the experiments presented in this paper. However, unlike the work by Evans and Wolfe, the capture effects observed in these experiments depended on the cued target on a specific trial. In fact, to the extent that unassociated contexts on a trial might have been associated with other targets, the capture effects described by Evans and Wolfe might mask the task-specific capture effects observed in the present experiments by increasing interference on trials with contexts unassociated with the target.

While these experiments were not designed to arbitrate among theories of the AB, they can address the filtering mechanisms that are increasingly central in recent accounts. Initially, the AB was believed to be the result of inhibitory mechanisms shielding the current contents of VSTM from competing distractors (Raymond et al., 1992). More recently, theoretical accounts of the AB focus on the role of attentional filters in the phenomenon. As mentioned previously, the temporary loss of control account argues that the attentional blink results from the reconfiguration of the attentional filter following an encounter with an initial target (DiLollo et al., 2005). The boost and bounce theory argues that the strong mismatch between a lag 1 distractor and the initial target results in an inhibitory signal that transiently disrupts processing at lags greater than 1 (Olivers, 2009). Regardless of the specific account of the AB, in most cases the evaluation of incoming stimuli is hypothesized to occur during a filtering stage where task relevant information is elaborated or consolidated. These experiments suggest that this filter is tuned quite broadly, selecting both the current target and

associated, but visually dissimilar, information. This is an important finding, because most studies of the attentional blink use relatively impoverished stimuli such as numbers, letters, or words.

Taken together, the experiments in this paper are consistent with conceptualizations of attentional control that emphasize high-level flexibility (Huettig, Olivers, & Hartsuiker, 2010; Olivers, 2010). Control of attention is flexible in the sense that observers are able to establish ACS for the identification of a verbally specified target quickly and consistently. Control of visual attention is argued to occur at a high level because the distractor-target relationship that resulted in costs was abstract and consisted of conceptual, associative linkages. Related contextual targets were visually dissimilar from targets. If attentional capture effects occur consistently on the basis of object-context relationship, this suggests that whatever criteria were employed for the selection of task-relevant objects is at least partly non-visual. Additionally, the observation of attentional capture with uniformly novel contextual distractors is inconsistent with any sensory or low-level account of the capture effects.

Attentional control, in the current experiments, seems to be operating on relatively elaborated representations of the presented objects and scenes. In terms of possible memory structures, this is consistent with accounts of volatile, but semantically elaborated, representations in CSTM (e.g. Potter, 1976). From the perspective of selection levels in attentional mechanisms, these data are broadly consistent with late selection accounts, where attention prioritizes task relevant stimuli only after the meaning of the stimulus is extracted (Deutsch &

Deutsch, 1963). Indeed, for reasons to be elaborated in the following section, there justification to conclude that the speed of object recognition is primarily of function of later stages of the ventral pathway (McKeeff, 2009). However, it should be mentioned that the locus of selection in classic cognitive psychological theories appears to depend critically on the perceptual load of task (Lavie, 1995). Therefore, any discussion of the selection stage in perceptual processing must be in the context of a particular task.

Implications for Theories of Object Recognition

These experiments not only have consequences for attentional control processes, but also for our understanding of object recognition more generally. In each novel experiment presented in this paper, observers detected or selectively encoded a verbally cued common object. There is an extensive research tradition that identifies the effects of visual context on object recognition sensitivity (e.g. Auckland et al., 2009) and bias (e.g. Hollingworth & Henderson, 1999). These present experiments extend this research tradition in several important ways. First, observers were required to identify a cued target. Because human object recognition is generally successful, observers are not typically provided with a label prior to the recognition task. The RSVP task in these experiments, with masking and brief presentations, was sensitive to contextual influences despite the fact that object category was known at the start of each trial. Second, the current experiments presented the associated context in a way that actually harms object recognition. This is important for two related reasons. On the one hand, in typical object recognition paradigms, observers'

responses regarding targets embedded in associated contexts are more accurate. Here they are less accurate. This shows that the effect of an associated context can be beneficial or detrimental depending on the temporal and spatial relationship between the context and target information sources. In much the same way that inaccurate spatial search cues can interfere with spatial visual search (Manginelli & Pollman, 2009, Fiske & Sanocki, 2010), inaccurate temporal cues can interfere with recognition during temporal visual search. In an additional consequence of this contextual cost, these experiments support the hypothesis that associated contexts automatically engage attentive processes. While this is important for our understanding of attentional control, this also has consequences for object recognition. Specifically, it suggests that searching for an object automatically activates representations of associated contexts. When items match these associated representations, they compete more successfully with targets that control stimuli.

These experiments replicate previous work indicating that certain contextual effects take time to accrue (e.g. Auckland et al., 2009). In Experiments 1, 2, and 3, observers did not show any effect of the contextual distractor at lag 1. Only after 80 - 100 ms, or one intervening distractor, did the contextual costs emerge. There are a number of hypothesized perceptual structures that might account for this delay. For example, certain interactivist accounts of visual cognition emphasize the role of re-entrant visual processes. These connections are argued to provide high-level hypotheses regarding earlier visual features. When there is a mismatch between the re-entrant and primary

representation, this disrupts processing, as is observed in the characteristic object substitution masking effect (DiLollo, Enns& Rensink, 2000). Bar provides another possible account that would predict the latency of this contextual effect (Bar, 2004). In this model, observers generate perceptual hypotheses about a scene based on low spatial frequency information available rapidly via magnocellular visual pathways. This coarse scene category information is then used by frontal areas to form hypotheses about the identity of individual objects as they are represented in the later stages of the ventral visual processing pathway. While these two models differ both in terms of content and process, they share the prediction that some contextual effects will depend critically on the relative timing of an object and a distractor.

Indeed, recent evidence suggests that temporal limitations in object recognition occur primarily due to factors late in object processing (McKeeff, 2009). McKeeff argues that neurons responsible for high-level object decisions, such as category membership, require a longer temporal window over which to integrate and analyze perceptual evidence for a given categorization. This longer temporal receptive field is analogous to the larger spatial receptive fields found anteriorly along visual processing pathways. Generally speaking, neurons early along the ventral visual pathway show greater location specificity and respond selectively to low-level visual attributes, such as color or orientation. In contrast, later neurons respond more robustly to high-level factors (e.g. face-ness) and show less location specificity. When multiple items appear in the same receptive field, information about about each item is diminished.

Desimone & Duncan (1995) elaborate this insight and present a broad framework for the integration of attention and working memory. Within the biased competition account of visual attention, objects compete for scarce cognitive resources. Evidence for this competition includes the interference effects observed when observers are required to respond to multiple simultaneously presented objects (Duncan, 1984). From a physiological perspective, the authors argue that, given a relatively fixed number and scope of receptive fields available in the ventral processing stream, the presence of additional objects in a given receptive field will decrease the available information regarding target objects. The outcome of this competition can be biased by either bottom-up or top-down factors. In the case of bottom-up factors, stimulus attributes including abrupt onsets or other transients will bias this system toward greater representation of those features associated with these salience-producing manipulations. Similarly, top-down factors such as the observer's current attentional set will favor the processing of some objects over others. Objects and features consistent with this set will be selectively enhanced at the expense of unrelated information. This biased competition account of visual attention has many advantages. It can treat object-centered behavioral effects in visual attention in a reasonably intelligible manner. By bringing visual attention and working memory into a shared theoretical framework, researchers can develop paradigms that treat broader scoped cognitive acts. The present experiments enhance this account by suggesting that competition within temporal receptive fields occurs on the basis of abstract knowledge.

Conclusions

Object recognition represents a ubiquitous and mysterious aspect of human mental life. Object recognition systems are implicated in virtually all aspects of complex human behavior, ranging from the mundane to the technical. Understanding and improving human performance in these tasks will require significant theoretical development. The present experiments describe the role of contextual associations in the selective encoding of verbally cued familiar objects. As such, they have implications for both our understanding of attentional control and models of object recognition. In terms of attentional control, these experiments describe tasks in which observers have high level control, but fail to be able to exclude clearly irrelevant, but conceptually related, object information. These data suggest that object recognition can be harmed by the presentation of associated contexts and relies on contextual information even in cases when it should not.

As is the case in so many perceptual domains, the very mechanisms that permit successful performance of complex tasks can limit performance in other cases. The associative knowledge about objects and scenes, that can support rapid and seemingly effortless object recognition in some settings can interfere with those very processes. Future research addressing these lapses in encoding control will have clear implications for both basic and applied question in visual cognition.

References

Adamo, M., Pun, C., Pratt, J., & Ferber, S. (2008). Your divided attention, please! The maintenance of multiple attentional control sets over distinct regions in space. *Cognition*, 107(1), 295-303.

Akyürek, E., & Hommel, B. (2006). Memory operations in rapid serial visual presentation. *European Journal of Cognitive Psychology*, 18(4), 520-536.

Anderson, A.K., Phelps, E.A. (2001). Lesions of the human amygdala impair enhanced perception of emotionally salient events. *Nature,* 411, 305-309.

Arguin, M., & Leek, E. C. (2003). Orientation invariance in visual object priming depends on prime-target asynchrony. *Perception and Psychophysics*, 65(3), 469-477.

Ariga, A, & Kawahara, J. I. (2004). The perceptual and cognitive distractor previewing effect. *Journal of Vision*, 4(10), 891-903.

Ariga, Atsunori, & Kawahara, J. I. (2004). The perceptual and cognitive distractor-previewing effect. *Journal of Vision*, 4(10), 891-903.

Asplund, C. L., Todd, J. J., Snyder, A. P., Gilbert, C. M., & Marois, R. (2010). Surprise-induced blindness: A stimulus-driven attentional limit to conscious perception.

Auckland, M. E., Cave, K. R., & Donnelly, N. (2007). Nontarget objects can influence perceptual processes during object recognition. *Psychonomic Bulletin and Review,* 14(2), 332-7.

Awh, E., Jonides, J., & Reuter-Lorenz, P. a. (1998). Rehearsal in spatial working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 24(3), 780-90.

Avrahami, J. (1999). Objects of attention, objects of perception. *Perception and Psychophysics*, 61(8), 1604-12.

Balcetis, E., & Dale, R. (2007). Conceptual set as a top-down constraint on visual object identification. *Perception*, 36(4), 581-95.

Bar, M., & Ullman, S. (1996). Spatial context in recognition. *Perception*, 25(3), 343-52.

Bar, M., & Aminoff, E. (2003). Cortical Analysis of Visual Context. *Neuron*, 38(2), 347-358.

Bar, M., Aminoff, E., & Schacter, D. L. (2008). Scenes unseen: the parahippocampal cortex intrinsically subserves contextual associations, not scenes or places per se. *The Journal of Neuroscience*, 28(34), 8539-44.

Barenholtz, E., & Tarr, M. J. (2007). Unsupervised learning of higher order statistics of visual features: evidence for relational encoding. *Journal of Vision*, 7(9), 798

Barenholtz, E., & Feldman, J. (2003). Visual comparisons within and between object parts: evidence for a single-part superiority effect. *Vision Research*, 43(15), 1655-1666. doi: 10.1016/S0042-6989(03)00166-4.

Barnard, P. J., Scott, S., Taylor, J., May, J., & Knightley, W. (2004). Paying attention to meaning. *Psychological Science*, 15(3), 179-86.

Baylis, G. C., & Driver, J. (1993). Visual attention and objects: Evidence for hierarchical coding of location. *Journal of Experimental Psychology Human Perception and Performance*, 19, 451

Becker, S. I. (2010). The role of target-distractor relationships in guiding attention and the eyes in visual search. *Journal of Experimental Psychology: General*, 139(2), 247-65.

Behrmann, M., Zemel, R. S., & Mozer, M. C. (1998). Object-based attention and occlusion: Evidence from normal participants and a computational model. *Journal of Experimental Psychology: Human Perception and Performance*, 24(4), 1011-1036.

Belke, E., & Humphreys, G.W., DG. (2008). Top-down effects of semantic knowledge in visual search are modulated by cognitive but not perceptual load. *Perception and Psychophysics*, 70(8), 1444-1458

Belopolsky, A. V., Schreij, D., & Theeuwes, J. (2010). What is top-down about contingent capture? Attention, *Perception, & Psychophysics*, 72(2), 326.

Biederman, I. (1981). On the semantics of a glance at a scene. Perceptual Organization, 213-253.

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: detecting and judging objects undergoing relational violations. *Cognitive Psychology,* 14(2), 143-77.

Boucart, M., & Bonnet, C. (1991). A study of the effect of structural information and familiarity in form perception. *The Quarterly Journal of Experimental Psychology: A, Human Experimental Psychology*, 43(2), 223-48. doi: 10.1080/14640749108400968.

Brady, T. F., & Chun, M. M. (2007). Spatial constraints on learning in visual search: Modeling contextual cuing. *Journal of Experimental Psychology: Human Perception and Performance*, 33(4), 798-815.

Breitmeyer, B G, & Ogmen, H. (2000). Recent models and findings in visual backward masking: A comparison, review, and update. *Perception and Psychophysics*, 62(8), 1572-1595.

Breitmeyer, B. G. (2007). Visual masking: past accomplishments, present status, future developments. *Advances in Cognitive Psychology* , 3(1-2), 9-20

Broadbent, D. E. (1958). Attention and communication. New York: Pergamon Press.

Brown, J. M., & Denney, H. I. (2007). Shifting attention into and out of objects: Evaluating the processes underlying the object advantage. *Perception and Psychophysics*, 69(4), 606.

Caetta, F., & Gorea, A. (2010). Upshifted decision criteria in attentional blink and repetition blindness. *Visual Cognition*, 18(3), 413-433.

Castelhano, M. S., & Heaven, C. (2010). The relative contribution of scene context and target features to visual search in scenes. *Attention, Perception, & Psychophysics*, 72(5), 1283-97.

Carrasco, M., Penpeci-Talgar, C., & Eckstein, M. (2000). Spatial covert attention increases contrast sensitivity across the CSF: support for signal enhancement. *Vision Research,* 40, 1203-1215.

Chaumon, M., Drouet, V., & Tallon-Baudry, C. (2008). Unconscious associative memory affects visual processing before 100 ms. *Journal of Vision*, 8(3), 10.1-10. doi: 10.1167/8.3.10.

Chun, M M. (1997). Types and tokens in visual processing: a double dissociation between the attentional blink and repetition blindness. *Journal of Experimental Psychology. Human Perception and Performance*, 23(3), 738-55.

Chun, M. M., & Jiang, Y. (2003). Implicit, long-term spatial contextual memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(2), 224-234.

Chun, M M, & Potter, M C. (1995). A two-stage model for multiple target detection in rapid serial visual presentation. *Journal of Experimental Psychology. Human Perception and Performance*, 21(1), 109-127.

Codispoti, M., Bradley, M. M., & Lang, P. J. (2001). Affective reactions to briefly presented pictures. *Psychophysiology*, 38(03), 474-478.

Codispoti, M., Ferrari, V., De Cesarei, A., & Cardinale, R. (2006). Implicit and explicit categorization of natural scenes. *Progress in Brain Research*, 156, 53-65.

Cowan, N., Fristoe, N. M., Elliott, E. M., Brunner, R. P., & Saults, J. S. (2006). Scope of attention, control of attention, and intelligence in children and adults. *Memory and Cognition,* 34(8), 1754-1768.

Crundall, D., Cole, G. G., & Galpin, A. (2007). Object-based attention is mediated by collinearity of targets. *Quarterly Journal of Experimental Psychology*, 60(1), 137-53.

Cuthbert, B. N., Schupp, H. T., Bradley, M., McManis, M., & Lang, P. J. (1998). Probing affective pictures: Attended startle and tone probes. *Psychophysiology*, 35(3), 344-347.

Davenport, J. L. (2007). Consistency effects between objects in scenes. *Memory and Cognition*, 35(3), 393-401.

Davenport, J. L., & Potter, M C. (2004). Scene consistency in object and background perception. *Psychological Science*, 15(8), 559-563.

Davis, G., Driver, J., Pavani, F., & Shepherd, A. (2000). Reappraising the apparent costs of attending to two separate visual objects. *Vision Research*, 40(10-12), 1323-32.

Davis, G., & Holmes, A. (2005). Reversal of object-based benefits in visual attention. *Visual Cognition*, 12(5), 817-846.

Deutsch, J. A., & Deutsch, D. (1963). Attention: Some Theoretical Considerations. *Psychological Review*, 70(1), 80-90.

DiLollo, V. D., & Enns, J. (2000). Competition for consciousness among visual events: The psychophysics of reentrant visual processes. *Journal of Experimental Psychology: General*, 129(4), 481-507.

Di Lollo, V., Enns, James T, & Rensink, Ronald A. (2000). Competition for consciousness among visual events: The psychophysics of reentrant visual processes. *Journal of Experimental Psychology: General*, 129(4), 481-507.

Downing, P. (2000). Interactions between visual working memory and selective attention. *Psychological Science*, 11(6), 467-473.

Draper, B. A, Collins, R. T., Brolio, J., Hanson, A. R., & Riseman, E. M. (1989). The schema system. *International Journal of Computer Vision*, 2(3), 209-250.

Dumais, S. T. (2004). Latent Semantic Analysis. *Annual Review of Information Science and Technology* (pp. 189-230).

Egly, R., Driver, J., & Rafal, R. D. (1994). Shifting visual attention between objects and locations: Evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology: General*, 123(2), 161-177.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, 16(1), 143-149.

Evans, K. K., & Treisman, Anne. (2005). Perception of objects in natural scenes: is it really attention free? *Journal of Experimental Psychology: Human Perception and Performance*, 31(6), 1476-1492.

Evans, K. K., & Wolfe, J. M. (2010). When Categories Collide ; Interference Effects in Gist Processing. Vision Sciences Society.

Fei-Fei, L., VanRullen, R., Koch, C., & Perona, P. (2005, August). Why does natural scene categorization require little attention? Exploring attentional requirements for natural and synthetic stimuli. *Visual Cognition*, 12(6), 893-924

Fendrich, R., Wessinger, M., & Gazzaniga, M. S. (2001). Speculations on the neural basis of blindsight. *Progress in Brain Research*, 134, 353-366.

Fiser, J., & Aslin, R. N. (2001). Unsupervised Statistical Learning of Higher-Order Spatial Structures from Visual Scenes. *Psychological Science*, 12(6), 499-504.

Fiser, J., & Aslin, R. N. (2005). Encoding multielement scenes: statistical learning of visual feature hierarchies. *Journal of Experimental Psychology: General*, 134(4), 521-537..

Fiske, S., & Sanocki, T. (2010). Memory and attentional guidance in contextual cueing. *Journal of Vision*, 10(7), 1302-1302.

Folk, C.L. & Gibson, B.S. (2001). Attraction, distraction and action : multiple perspectives on attentional capture. *Advances in Psychology,*133

Folk, C.L., Leber, A. B., & Egeth, H. E. (2007). Top-down control settings and the attentional blink: Evidence for nonspatial contingent capture. *Visual Cognition*,16(5), 616-642.

Folk, C.L., Remington, R.W. (2008). Selectivity in distraction by irrelevant featural singletons: Evidence for two forms of attentional capture. *Journal of Experimental Psychology: Human Perception and Performance*, Vol 24(3)

Folk, C L, Remington, R. W., & Johnston, J C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance*, 18(4), 1030-44.

Folk, C.L., Remington, R.W., & Wright, J.H. (1994). structure of attentional control: Contingent attentional capture by apparent motion, abrupt onset, and color. *Journal of Experimental Psychology: Human Perception and Performance*, Vol 20(2)

Forti, S., & Humphreys, Glyn W. (2008). Sensitivity to object viewpoint and action instructions during search for targets in the lower visual field. *Psychological Science*, 19(1), 42-48.

Gaillard, R., Del Cul, A., Naccache, L., Vinckier, F., Cohen, L. & Dehaene, S. (2006). Nonconscious semantic processing of emotional words modulates conscious access. *Proceedings of the National Academy of Sciences,* 103(19), 7524-7529.

Ghorashi, S., Enns, J T, Klein, R. M., & Di Lollo, V. (2003). Spatial selection and target identification are separable processes in visual search. *Journal of Vision*, 10(3), 7.1-12.

Goujon, A., Didierjean, A., & Marmèche, E. (2007). Contextual cueing based on specific and categorical properties of the environment. *Visual Cognition*, 15(3), 257-275.

Graham, N. (1985). Detection and identification of near-threshold visual patterns. *Journal of the Optical Society of America,* 2, 1468-1482

Gratton, G., Coles, M. G., Sirevaag, E. J., Eriksen, C. W., & Donchin, E. (1988). Pre- and poststimulus activation of response channels: a psychophysiological analysis. *Journal of Experimental Psychology. Human Perception and Performance*, 14(3), 331-44.

Greene, M. R., & Oliva, Aude. (2008). Recognition of natural scenes from global properties: seeing the forest without representing the trees. *Cognitive Psychology*, 58(2), 137-76.

Greene, M. R., & Oliva, Aude. (2009). The Briefest of Glances: The Time Course of Natural Scene Understanding. *Psychological Science*, 20(4), 464-472.

Grider, R. C., & Malmberg, K. J. (2008). Discriminating between changes in bias and changes in accuracy for recognition memory of emotional stimuli. *Memory & Cognition*, 36(5), 933-946.

Hayward, W. G., & Tarr, M. J. (1997). Testing conditions for viewpoint invariance in object recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 23(5), 1511-1521.

Henderson, J. M., & Ferreira, F. (2004). Scene Perception for Psycholinguists. The interface of language, vision, and action: Eye movements and the visual world. In J. M. Henderson (Ed.), The interface of language, vision, and action: Eye movements and the visual world. New York, NY : Psychology Press.

Hoffman, D. D. (1984). Parts of recognition. *Cognition*, 18(1-3), 65-96. doi: 10.1016/0010-0277(84)90022-2.

Hollingworth, a, & Henderson, J M. (1998). Does consistent scene context facilitate object perception? *Journal of Experimental Psychology: General*, 127(4), 398-415.

Hollingworth, A, & Henderson, John M. (1999). Object identification is isolated from scene semantic constraint: evidence from object type and token discrimination. *Acta Psychologica*, 102(2-3), 319-343.

Hollingworth, Andrew, & Henderson, J.M. (2000). Semantic informativeness mediates the detection of changes in natural scenes. *Visual Cognition*, 7(1), 213-235.

Hollingworth, Andrew, & Henderson, J.M. (2003). Testing a conceptual locus for the inconsistent object change detection advantage in real-world scenes. *Memory and Cognition*, 31(6), 930-940.

Humphreys, G. W., & Heinke, D. (1998). Spatial Representation and Selection in the Brain: Neuropsychological and Computational Constraints. *Visual Cognition*, 5(1), 9-47.

Huettig, F., Olivers, Christian N.L., & Hartsuiker, R. J. (2010). Looking, language, and memory: Bridging research from the visual world and visual search paradigms. *Acta Psychologica*, 1-13.

Intraub, H. (1984). Conceptual masking: the effects of subsequent visual events on memory for pictures. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10(1), 115-125.

Intriligator, J., & Cavanagh, P. (2001). The Spatial Resolution of Visual Attention. *Cognitive Psychology*, 43(3), 171-216.

Itti, L., & Arbib, M. A. (2005). Attention and the Minimal Subscene. Action to Language via the Mirror Neuron System.

Joubert, O. R., Fize, D., Rousselet, G. A., & Fabre-Thorpe, M. (2008). Early interference of context congruence on object processing in rapid visual categorization of natural scenes. *Journal of Vision*, 8(13), 1-18.

Kahneman, D., & Treisman, A. (1984). Changing views of attention and automaticity. *Varieties of Attention*, 29-61.

Kanwisher, N. G., Kim, J. W., & Wickens, T. D. (1996). Signal detection analyses of repetition blindness. *Journal of Experimental Psychology: Human Perception and Performance*, 22(5), 1249-1260.

Kanwisher, N. G., & Potter, M C. (1990). Repetition blindness: levels of processing. *Journal of Experimental Psychology: Human Perception and Performance*, 16(1), 30-47.

Karklin, Y. & Lewicki, M. S. (2003). Learning higher-order structures in natural images. *Network: Computation in Neural Systems*, 14(3), 483-499

Koivisto, M., & Revonsuo, A. (2007). How meaning shapes seeing. *Psychological Science*, 18(10), 845-9.

Kunar, M. A., Flusberg, S., Horowitz, T. S., & Wolfe, J. M. (2007). Does contextual cuing guide the deployment of attention? *Journal of Experimental Psychology: Human Perception and Performance*, 33(4), 816-28.

LaBerge, D. (2002). Attentional control: brief and prolonged. *Psychological Research*, 66(4), 220-33.

Lang, P. J., Greenwald, M. K., Bradley, M. M., & Hamm, A. O. (1993). Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology*, 30(3), 261-273.

Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human Perception and Performance*, 21(3), 451-468.

Leber, A. (2004). On the maintenance and switching of attentional set. Dissertation Abstracts International: Section B: The Sciences and Engineering. 2004, pp. 5247

Leber, A. B., & Egeth, H. E. (2006). It's under control: Top-down search strategies can override attentional capture. *Psychonomic Bulletin and Review*, 13(1), 132.

Leber, A. B., Kawahara, J. I., & Gabari, Y. (2009). Long-term abstract learning of attentional set. *Journal of Experimental Psychology: Human Perception and Performance*, 35(5), 1385-1397.

Lewis, J. L. (1970). Semantic processing of unattended messages using dichotic listening. *Journal of Experimental Psychology*, 85(2), 225-228.

Lien, M.-C., Ruthruff, E., & Johnston, James C. (2010). Attentional capture with rapidly changing attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance*, 36(1), 1-16.

Logan, G. D. (1992). Attention and preattention in theories of automaticity. *American Journal of Psychology*, 105(2), 317-339.

Luck, S., Vogel, E., & Shapiro, K. (1996). Word meanings can be accessed but not reported during the attentional blink. *Nature*, 383, 616-618.

Mack, M. L., & Palmeri, T. J. (2010). Decoupling object detection and categorization. *Journal of Experimental Psychology: Human Perception and Performance*, 36(5), 1067-1079.

Mack, A., & Rock, I. (1998). Inattention blindness. Cambridge, MA: MIT Press.

Mackay, D. G. (1973). Aspects of the theory of comprehension, memory and attention. *The Quarterly Journal of Experimental Psychology*, 25(1), 22-40.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, 109(2), 163-203.

Malcolm, G. L., & Henderson, J.M. (2010). Combining top-down processes to guide eye movements during real-world scene search. *Journal of Vision*, 10(2), 4.1-11.

Maljkovic, V., & Nakayama, K. (1994). Priming of pop-out: 1. Role of Features. *Memory and Cognition*, 22(6), 657–672.

Manginelli, A. A., & Pollmann, S. (2009). Misleading contextual cues: How do they affect visual search? *Psychological Research*, 73(2), 212-221.

Maki, W.S. & Mebane, M.W. (2006). Attentional capture triggers an attentional blink. *Psychonomic Bulletin and Review*, 13, 1, 125-131

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review*, 88(5), 375-407.

Mckeeff, T. J. (2009). Temporal Limitations of Visual Object Processing. Dissertation Abstracts International: Section B: The Sciences and Engineering

Moore, C. M., Hein, E., Grosjean, M., & Rinkenauer, G. (2009). Limited influence of perceptual organization on the precision of attentional control. *Attention, Perception, & Psychophysics*, 71(4), 971

Moray, N. (1959). Attention in dichotic listening: Affective cues and the influence of instructions. *The Quarterly Journal of Experimental Psychology*, 11 (1), 56 - 60

Moray, N. (1970). Attention: Selective processes in vision and hearing. Hutchinson.

Most, S. B., Chun, Marvin M, Widders, D. M., & Zald, D. H. (2005). Attentional rubbernecking: cognitive control and personality in emotion-induced blindness. *Psychonomic Bulletin and Review*, 12(4), 654-661.

Most, S. B., Laurenceau, J.P., Graber, E., Belcher, A., & Smith, C.V. (2010). Blind jealousy? Romantic insecurity increases emotion-induced failures of visual perception. *Emotion*, 10, 250-256.

Most, S.B., Simons, D.J., Scholl, B.J., Chabris, C.F. (2000). Sustained inattentional blindness. *Psyche*, 6, 14

Nakayama, K., & Martini, P. (in press). Situating visual search. *Vision Research*. Elsevier.

Navalpakkam, V., & Itti, L. (2003). A Goal Oriented Attention Guidance Model. BMCV03.

Norman, D. (1975). On data-limited and resource-limited processes*1. *Cognitive Psychology*, 7(1), 44-64.

Oliva, A, Torralba, A., Castelhano, M., & Henderson, J.M. (2003). Top-down control of visual attention in object detection. *Proceedings 2003 International Conference on Image Processing* (pp. I-253-6). IEEE.

Olivers, Christian N L. (2010). Long-term visual associations affect attentional guidance. *Acta psychologica*, 1-5.

Palmer, S. E. (1975). The effects of contextual scenes on the identification of objects. *Memory & Cognition*, 3(5), 519-526.

Pashler, H.E. (1999) The psychology of attention. MIT Press, Cambridge, MA.

Pelli, D. G., & Farell, B. (1995). Psychophysical Methods. Handbook of Optics. McGraw-Hill.

Phelps, E., Ling, S., Carrasco, M. (2006). Emotional facilitates perception and potentiates the perceptual benefits of attention. *Psychological Science,* 17, 292-299.

Posner, M. I. (1980). Orienting of attention. *The Quarterly Journal of Experimental Psychology*, 32(1), 3-25.

Posner, M. I., & Cohen, Y. (1984). Components of visual orienting. In H. Bouma, BouwheisD.G. (Ed.), *Attention and performance X: Control of language processes* (32), 531-554.

Posner, M. I., Rafal, R. D., Choate, L. S., & Vaughan, J. (1985). Inhibition of return: Neural basis and function. *Cognitive Neuropsychology*, 2(3), 211-228.

Potter, M. C. (1975). Meaning in visual search. *Science*, 187(4180), 965-966.

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory*, 2(5), 509-22.

Potter, M. C. (1999). Understanding sentences and scenes: The role of conceptual short-term memory. *Fleeting memories: Cognition of brief visual stimuli*, 13-46.

Potter, M. C.. (1975). Meaning in Visual Search. *Science*, 187(4180), 965-966.

Potter, M. C., Staub, A., Rado, J.. (2002). Recognition memory for briefly presented pictures: The time course of rapid forgetting. *Journal of Experimental Psychology: General*, 28(5), 1163-1175.

Pylyshyn, Z. (2001, June). Visual indexes, preconceptual objects, and situated vision. Cognition, 80(1), 127-158.

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: evidence for a parallel tracking mechanism. *Spatial Vision*, 3(3), 179-197.

Rensink, R. A. (2000). The dynamic representation of scenes. *Visual Cognition*, 7(1), 17-42.

Rensink, R. a, O'Regan, J. K., & Clark, J. J. (1997). To See or not to See: The Need for Attention to Perceive Changes in Scenes. *Psychological Science*, 8(5), 368-373.

Robertson, L., Treisman, Anne, Friedman-Hill, S., & Grabowecky, M. (1997). The Interaction of Spatial and Object Pathways: Evidence from Balint's Syndrome. *Journal of Cognitive Neuroscience*, 9(3), 295-317.

Rosa, S. de la, Choudhery, R. N., & Chatziastros, A. (2011). Visual object detection, categorization, and identification tasks are associated with different time courses and sensitivities. *Journal of Experimental Psychology: Human Perception and Performance,* 37(1), 38-47.

Rubin, E. (1915). Visuell Wahrgenommene Figure. Copenhagen: Gyldenalske Boghandel.

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: II. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review*, 89(1), 60-94.

Sanocki, T., Bowyer, K., Heath, M., & Sarka, S. (1998). Are edges sufficient for object recognition? *Journal of Experimental Psychology: Human Perception and Performance*, 24(1), 340-349.

Sanocki, T., & Oden, G. C. (1984). Contextual validity and the effects of low-constraint sentence contexts on lexical decisions. *The Quarterly Journal of Experimental Psychology. A. Human Experimental Psychology*, 36(1), 145-156.

Schankin, A., & Schubö, A. (2010). Contextual cueing effects despite spatially cued target locations. *Psychophysiology*, 47(4), 717-27.

Scholl, B. J. (2001). Objects and attention: the state of the art. *Cognition,* 80(1-2), 1-46.

Schmukle, S. C. (2005). Unreliability of the Dot Probe Task. *European Journal of Personality*, 595-605.

Sestokas, A. K., Lehmkuhle, S., and Kratz, K. E. (1987). Visual latency of ganglion X- and Y-cells: A comparison with geniculate X- and Y-cells. *Vision Research*, 27, 1399-1408.

Schmolesky MT, Wang YC, Hanes DP, Thompson KG, Leutgeb S, Schall JD, Leventhal, AG (1998) Signal timing across the macaque visual system. J *Neurophysiology*, 79: 3272-3278.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, 84(2), 127-190.

Shinoda, H., Hayhoe, M. M., & Shrivastava, A. (2001). What controls attention in natural environments? *Vision Research*, 41(25-26), 3535-3545.

Shomstein, S., & Behrmann, M. (2008). Object-based attention: Strength of object representation and attentional guidance. *Perception and Psychophysics*, 70(1), 132-144.

Shomstein, S., & Behrmann, M. (2008). Object-based attention: Strength of object representation and attentional guidance. *Perception & Psychophysics*, 70(1), 132-144.

Shomstein, S., & Yantis, S. (2002). Object-based attention: Sensory modulation or priority setting? *Perception & Psychophysics*, 64(1), 44.

Simons, D J, & Rensink, R A. (2005). Change blindness: Past, present, and future. *Trends in Cognitive Sciences*, 9(1), 16-20. Elsevier.

Tanaka, K. (2002). Neuronal representation of object images and effects of learning. Perceptual learning (pp. 67-82). Cambridge, MA US: MIT Press.

Theeuwes, J. (1994). Stimulus-driven capture and attentional set: selective search for color and visual abrupt onsets. *Journal of Experimental Psychology. Human Perception and Performance*, 20(4), 799-806.

Theeuwes, J. (2004). Top-down search strategies cannot override attentional capture. Psychonomic Bulletin and Review, 11(1), 65.

Theeuwes, J., Kramer, A. F., Hahn, S., Irwin, D. E., & Zelinsky, G.J. (1999). Influence of attentional capture on oculomotor control. *Journal of Experimental Psychology: Human Perception and Performance*, 25(6), 1595–1608.

Tipper, S. P., & Weaver, B. (1998). The medium of attention: Location-based, object-centered, or scene-based? In R. D. Wright (Ed.), Visual attention, Vancouver studies in cognitive science (p. 478). Oxford University Press.

Treisman, A. M. (1960). Contextual cues in selective listening. *The Quarterly Journal of Experimental Psychology*, 12(4), 242-248.

Treisman, A. M. (1964). The effect of irrelevant material on the efficiency of selective listening. *The American Journal of Psychology*, 77(4), 533-546.

Treisman, A, & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review*, 95(1), 15-48.

Vecera, S. P., & Farah, M. J. (1994). Does visual attention select objects or locations? *Journal of Experimental Psychology: General*, 123(2), 146-160.

Vickery, T., King, L. W., & Jiang, Y. V. (2005). Setting up the target template in visual search. *Journal of Vision*, 5(1), 81-92.

Watson, D. G., & Humphreys, G. W. (2000). Visual marking: Evidence for inhibition using a probe-dot detection paradigm. *Perception and Psychophysics*, 62(3), 471-481

White, R. C., & Aimola Davies, A. (2008). Attention set for number: expectation and perceptual load in inattentional blindness. *Journal of Experimental Psychology. Human Perception and Performance*, 34(5), 1092-1107.

Wolfe, J. M. (1998). What Can 1 Million Trials Tell Us About Visual Search? *Psychological Science*, 9(1), 33-39.

Wolfe, J. M. (2006). Guided Search 4.0 Current Progress With a Model of Visual Search. Search, 99-120.

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 15(3), 419-433.

Yantis, S., & Jonides, J. (1996). Attentional capture by abrupt onsets: New perceptual objects or visual masking? *Journal of Experimental Psychology-Human Perception and Performance*, 22(6), 1505-1513.

Zeelenberg, R., Wagenmakers, E.-J., & Rotteveel, M. (2006). The impact of emotion on perception: bias or enhanced processing? *Psychological Science*, 17(4), 287-91.

Appendices

Appendix 1

Table A1: List of Object, Context Categories

Object Category	Associated Context
airplane	airport
alarm	bedside table
ambulance	hospital
apple	produce department
armchair	living room
barbell	gym
baseball	baseball diamond
basketball	basketball court
beer bottle	cooler
bike	bike rack
camel	desert
car	freeway
cash register	checkout
church bell	wedding interior
circular saw	workshop
clothing iron	ironing board
computer	computer desk
cookie	cookie jar
cooking pan	stovetop
cow	farm
cowboy boots	cowboy
crane	construction site
dog	doghouse
doll	dollhouse
duck	pond
elephant	zoo
fence	yard
football	football field
goldfish	aquarium
grill	patio
hairdryer	hair salon
hammer	toolbox
handcuffs	police
hockey stick	hockey rink

Appendix 1 (Continued)

Table A1 (Continued)

horse	farm
jack in the box	toy box
lunch tray	cafeteria
money bag	vault
necklace	jewelry box
oar	rowboat
paint palette	art studio
pancakes	breakfast
pepperoni	pizza
pillow	bed
robin	nest
saddle	horse context
sailboat	lake
scuba	underwater
seagull	beach
seashell	beach
shopping cart	grocery store
skis	ski lodge
soccer ball	soccer game
spider	spider web
stethoscope	doctor's office
swan	pond
table knife	dish rack
tennis racquet	tennis court
tent	camp site
toaster	kitchen
toilet	bathroom
toothbrush	toothpaste counter
tractor	farm
train	railroad tracks
violin	orchestra
volleyball	volleyball court
watering can	garden
wedding ring	wedding ceremony
wrapped present	christmas tree

Appendix 1 (Continued)

Table A2 Object Images

Appendix 1 (Continued)

Table A2 Object Images (Continued)

Appendix 1 (Continued)

Table A3 Context Images

Appendix 1 (Continued)

Table A3 Context Images (Continued)