INFLUENCES OF CONTEXT ON OBJECT DETECTION AND IDENTIFICATION

INFLUENCES OF CONTEXT ON OBJECT DETECTION AND IDENTIFICATION IN NATURAL SCENES

By MITCHELL R. P. LAPOINTE, B.A. (Hon), M.Sc.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy

McMaster University

DOCTOR OF PHILOSOPHY (2016) McMaster University (Psychology)

TITLE: Influences of Context on Object Detection and Identification in Natural Scenes

AUTHOR: Mitchell R. P. LaPointe, B.A. (Hon) (St. Thomas University), M.Sc. (University of Lethbridge)

SUPERVISOR: Professor Bruce Milliken

NUMBER OF PAGES: xviii, 174

Lay Abstract

The way that we perceive and create mental representations of our visual world has been an area of debate in cognitive psychology. The research attempting to address these issues has reported contradictory findings. For example, some studies have shown that the context of a scene is important for efficient perception of that scene and its parts. Other studies, however, have shown that the context of a scene can undermine efficient perception of scenes. The current research identifies two distinct processes that underlie complex scene perception. One process appears to benefit from a congruent context, whereas the other appears to benefit from an incongruent context. Further, the weighting of these two processes can produce context congruency benefits in some experimental situations and congruency costs in others. Finally, it appears as though when processing is weighted towards congruency costs, attention is attracted to incongruent objects early into scene processing.

Abstract

The way we perceive complex visual scenes has been an area of much research and debate. Many studies have found that the context of a scene is used to guide attention to important and relevant areas of a scene. Other studies, however, have found that objects that are incongruent with the scene context capture attention. These contradictory findings have been found both within and across tasks. The purpose of the present research was to reconcile these contradictory results. Two processes were identified as underlying complex scene perception: object detection and object identification. Further, the current research demonstrates the relative weighting of these processes differs according to task demands; some tasks weight object detection more heavily, whereas other tasks weight object identification more heavily. Moreover, it was demonstrated that the weighting of these processes can be manipulated within a task in such a way as to produce either congruency benefits or costs. Finally, in circumstances in which processing is weighted in favour of object detection, it was demonstrated that eye gaze, and presumably overt attention, is captured by semantically incongruent objects early into scene perception.

The current research helps our understanding of complex scene perception by reconciling contradictory findings reported in previous studies. In particular, two processes were identified: object identification, which relies on a congruent context, and object detection, which relies on an incongruent context. In this way, past experience may promote efficient scene perception by promoting the use of regularities in the environment

(e.g., congruent context), but also leaving the attention system sensitive to areas of the scene that contradict the expectations set by the context (e.g., incongruent objects).

Acknowledgements

This thesis is the product of support from a large community. Without this support the current research could not have been completed. I am grateful for my time doing graduate work at McMaster University and the many friendships that I have made. In particular, the Department of Psychology, Neuroscience & Behaviour has been an inspiring and nurturing place for a new researcher to grow. The atmosphere is thanks to years of tradition that encourage collaboration and socialization. Through reading groups, talks, lab meetings, and conversations with colleagues, my thought process has routinely been taken to new and exciting places. For this, I am grateful. There are a few in this supportive community that deserve a special note of gratitude.

First and foremost, I owe a large debt of gratitude to my supervisor, Bruce Milliken. Thank you, Bruce, for being a kind and patient teacher. Your passion for research is infectious. I am grateful for your attention to detail and for pushing me to always be better. Every aspect of this thesis, from conceptual issues to writing, is vastly improved thanks to your prodding. Thank you for your loyalty. To a young researcher, confidence can be fleeting, but knowing your primary supervisor supports you and your research makes success all the more tenable. The lessons in experimental psychology you have passed on will stay with me.

To my other co-author, Juan Lupiáñez, thank you for your insightful comments and your excitement as we pursued the research project contained in the second chapter. Thank

you for sharing your graduate students with us, and for graciously hosting our lab in Granada, where we discussed our mutual interests in cognitive psychology.

To my committee members, Judy Shedden and Hongjin Sun, thank you for your support and encouragement. Your comments and ideas throughout this process have always been insightful, thoughtful, and encouraging. Thank you for directing me to articles and areas of research I had not considered. Your input has fundamentally shaped this research and my thinking more broadly.

I also owe a special thank you to my lab mates, past and present. Thank you to Sandra Monteiro, Chris Fiacconi, Maria D'Angelo, Dave Thomson, Adam Spadaro, and Ellen MacLellan for introducing me to the lab and embracing me as a colleague and friend. Sandra, my first office mate, I appreciate our conversations about different lines of research and how they may converge in one way or another. Chris, your philosophical way of thinking about issues is contagious; I have always enjoyed our chats. Maria, more than one research idea contained in these pages is due directly to a comment you made in our lab meeting or in passing, thank you. Dave and Adam, thank you both for your friendship, our conversations about research and beyond, and our social time. You both made my time as a graduate student a very happy experience. Ellen, the social glue that makes our lab such a cohesive and friendly group, thank you for always listening to me rant about research and other things. You are one of the most selfless people I have ever met. Thank you for being such a great friend.

Thank you to my current lab mates, Tammy Rosner, Robert Collins, Brett Cochrane, Hanae Davis, and Lisa Lorentz. As the personnel has changed, I have been delighted that the lab tradition of collaboration and support has continued. Tammy, thank you for bringing a sunny demeanour to the lab and for your constant encouragement. Robert, I've appreciated our research and political conversations. Brett, I appreciate your inquisitive nature; your questions have made my research and communication better. Hanae, thank you for always pushing me and others conceptually forward. Lisa, thank you for all of your hard work on the projects we have worked on together.

Beyond academic support, this research could not have been completed without the support of family and friends. Courtnay, when my experiments did not turn out the way I expected, when the data left me perplexed, and when the rigours of the doctoral program made me anxious, you kept me grounded with your unconditional love and support. Thank you for suffering through my research 'stories', enduring my absent-mindedness, and for your willingness to entertain research ideas, even at the most inopportune times. You make life more enjoyable. I love you.

To my parents, Joel and Elaine, thank you for instilling in me a curiosity about the world, and for your patience as my siblings and I took over family dinners with light-hearted argument. Thank you for always supporting me. From the day I jumped in my car to drive across the country for graduate school, you have encouraged me. I could not have finished this project without your unconditional love and support. Thank you.

Table of Contents

List of Figures

CHAPTER 2

List of Tables

Preface

The following is a 'sandwich' thesis. The first two empirical chapters (Chapters 2 and 3) have been published in peer-reviewed journals. The third empirical chapter (Chapter 4) has been submitted for publication and is currently under peer review. I am the first author for each of the empirical chapters, my supervisor is a co-author for each of these chapters, and a collaborator is a co-author on the first empirical chapter. The following paragraphs detail my contributions to each of the empirical chapters.

The first empirical chapter is a reprint of LaPointe, M. R. P., Lupiáñez, J., & Milliken, B. (2013). Context congruency effects in change detection: Opposing effects on detection and identification. *Visual Cognition, 21*, 99-122. I was involved in all aspects of the research, including conceptual development, experimental design, computer programming, data collection, analyses, and I was the primary writer.

The second empirical chapter is a reprint of LaPointe, M. R. P., & Milliken, B. (in press). Semantically incongruent objects attract eye-gaze when viewing scenes for change. *Visual Cognition*, Online first publication. http://dx.doi.org/ 10.1080/13506285.2016.1185070. For this chapter, I was involved in conceptual development, experimental design, analyses, and I was the primary writer.

The third and final empirical chapter is a reprint of LaPointe, M. R. P., & Milliken, B. (in review). Conflicting effects of context in change detection and visual search: A dual process account. *Canadian Journal of Experimental Psychology.* Manuscript ID:

CEP-2016-1316. I was involved in conceptual development, experimental design, computer programming, data collection, analyses, and I was the primary writer.

Each of the empirical chapters were originally written as stand alone scientific articles. As such, there is some redundancy in the introduction and general discussion sections across chapters in this thesis. Regardless, each empirical chapter and the experiments within were designed to address separate theoretical issues.

Chapter 1: Introduction

The visual world is complex. At any given moment, a number of objects may be located within the visual field, with each object made up of numerous features, such as line edges, contours, orientations, colours, luminance intensities, and moving parts. Creating and maintaining mental representations of the visual world is a daunting task given this complexity. Indeed, there is a large literature on visual perception that shows that our representations of the visual world are often incomplete, not very robust, and created in a piecemeal fashion. These scientific observations about vision are particularly interesting because they contradict our everyday impression that vision offers a robust and veridical representation of the world around us.

A powerful demonstration of just how incomplete our visual representations can be comes from studies on inattentional blindness (Mack & Rock, 1998). One method used to study inattentional blindness presents participants with a dynamic visual display and with the task being to monitor one aspect of that display. When experimenters subsequently probe participants about a secondary aspect of the display, participants often show no awareness. The classic demonstration of this effect comes from an experiment by Neisser and Becklen (1975), but has since been replicated and extended by Simons and Chabris (1999). In the modern version of the experiment, participants are shown a video of two groups of people, one group wearing black shirts and the other wearing white shirts, with each group passing around a basketball. Participants are told that they must maintain a mental count of each time those wearing either the white or black shirts pass the basketball.

Part way through the video, a person carrying an umbrella or wearing a gorilla costume walks past the two groups in full view of the camera. At the end of the video, when asked about what they saw, collapsing across all conditions, only about one third of participants report having seen the unexpected event.

This phenomenon has been called inattentional blindness because although the unexpected event (e.g., person in a gorilla suit) occurred in full view of the camera and landed on the retina of the viewer, attention was preoccupied with another event, thereby leaving the viewer "blind" to the unexpected event. Interestingly, participants were more apt to report noticing the person carrying the umbrella walking across the scene than the person wearing a gorilla suit. Simons and Chabris (1999) argued that this may occur because a person, rather than a gorilla, is more in line with expectations of scenes with people passing basketballs. Further, participants were more apt to notice the person in the gorilla suit when their task was to monitor those wearing black shirts, presumably because a basic physical property of the unexpected event (i.e., colour) matched a basic physical property of the attended events.

Further demonstrations of incomplete representations of the visual world come from the change blindness literature (Rensink, O'Regan, & Clark, 1997; Simons & Levin, 1997; Simons & Rensink, 2005). In a common task used here, a picture is presented briefly and then presented again with one aspect of the scene changed or missing. A crucial aspect of this task is a blank display (usually, but not necessarily white) presented briefly between the picture presentations. This blank display is argued to affect visual input in much the

same way as eye saccades and eye blinks that occur during natural viewing conditions (Rensink et al., 1997). More important, the blank display disrupts fluid viewing and processing of the pictures. Under these conditions, large and often important aspects of the scene can be changed from one presentation to the next without the viewer being aware. These findings, and those coming from studies on inattentional blindness, demonstrate that our representations of the visual world are not complete, veridical, or seamless. From one moment to the next, there is visual information that is not consciously registered by the viewer, and this processing deficit can have significant consequences for behaviour. Moreover, these studies demonstrate that attention is selective and is required when extracting information from a visual scene (see Moray, 1959 for an example of the selective nature of attention in the auditory modality).

If attention is selective when viewing the world, what processes underlie this selection? Put another way, what drives the allocation of attention to one aspect of a visual scene over another? This question has been the subject of a significant amount of study and debate, however, researchers generally agree that selectivity of attention can be divided into two broad categories: exogenous and endogenous (see Ruz & Lupiáñez, 2002 for a review). Exogenous attention refers to processes by which the environment captures attention automatically, without the need for limited capacity cognitive resources. For example, when a stimulus is particularly striking, attention is drawn to it without effort or intention on the part of the viewer. In contrast, endogenous attention refers to slow, effortful allocation of attention that is under the control of the viewer. For example, while

looking for a lost set of car keys, it is best to actively and strategically guide attention to areas in which the keys are likely to be found, rather than passively letting salient stimulus features guide attention.

Selection Models of Attention

Whether attention is captured exogenously or allocated endogenously, Broadbent (1958) proposed that all information must first be processed at a physical feature level before meaning can be extracted. This idea has been referred to as the early selection model of attention, according to which attention is required to access semantic information. In other words, the semantic properties of a stimulus (e.g., its identity) cannot capture attention. Broadbent argues that the processing of physical features (e.g., line edges, contours, luminance) can occur in parallel for multiple items. The representations of these features are then stored in a peripheral memory store that has a high capacity. The stage beyond the peripheral store, at which semantic information is accessed, is of limited capacity. As such, a selective filter is required to ensure that items are processed serially at this stage. The selective filter is used to select relevant information, as well as to filter out unwanted information. Once the selected information is adequately processed semantic information becomes available, and the representation can then be integrated with information in a long-term memory store.

Some researchers, however, have pointed to evidence that seems to contradict Broadbent's (1958) view that attention is required to extract meaning. A classic example of such evidence is the tendency for listeners to prioritize their own name over information

that is currently the focus of attention (Morey, 1959). In this example, the participant is exposed to two streams of auditory information and is asked to selectively attend to one of the two streams while ignoring the other. If, however, the participant's name is presented as part of the ignored auditory stream, then attention is captured, overriding selection of the primary stream. This result would seem to contradict Broadbent's assertion that attention is required to extract meaning from a stimulus. In this case, it appears that the meaning contained in the name is capturing attention. In part due to results such as this one, a late selection model of attention has been proposed, whereby attention selects relevant information not only on the basis of physical features of stimuli, but also on the basis of intention, motivation, and past experience of the actor (Deutsch & Deutsch, 1963). From this perspective, in some circumstances semantic information is processed in the absence of attention, and therefore may serve as a cue to capture attention.

Attention Capture by Perceptual Features

Much of the work over the past half-century has supported a key tenet of early selection models of attention, in particular that physical features of stimuli can capture attention while higher-level semantic characteristics cannot. This idea is well captured by feature-integration theory (Treisman, Sykes, & Gelade, 1977; Treisman & Gelade, 1980). According to this theory, simple physical features of objects are processed early, automatically, and in parallel across a visual scene. Integrating the features, however, requires the focus of attention, which is allocated in a serial fashion from one area of the

scene to the next. Once attention has been dedicated to an object and its features are integrated, the whole object can be stored in memory.

To test the assumptions of feature-integration theory, Treisman and Gelade (1980) conducted a series of visual search experiments. In one experiment, participants were presented visual arrays consisting of 1, 5, 15, or 30 simple objects. Each visual array was made up of a group of distractor items that consisted of 'T' shapes presented in brown and 'X' shapes presented in green. In the conjunction condition, target objects consisted of 'T' shapes presented in green. Note, the target object in this condition shares a feature with both distractor items and, thus, to correctly identify the target both features (i.e., colour and shape) must be processed in conjunction to differentiate it from the distractor items. In the feature condition, the target object could be either a 'T' or 'X' presented in blue, or an 'S' presented in either brown or green. Note, the target objects in this condition share a feature with the distractor items, but also contain a feature independent of the distractor items, and thus only one feature (i.e., colour on some trials, shape on the other trials) needs to be processed to differentiate the target from the distractors. In both conditions, participants were to indicate whether a target object was present or absent from the display.

To assess attention capture in visual search tasks, response times for correct trials are compared across conditions that vary in the number of distractors in the display. If response times increase as the number of distractor items increase, attention is argued to be allocated in a serial fashion from one object to the next and is taken as an indication that the target object does not capture attention automatically. In contrast, if response times

remain relatively stable as the number of distractor items increase, attention is argued to be allocated in parallel across the visual display with the target object capturing attention automatically.

Treisman and Gelade (1980) reasoned that if targets defined by simple features capture attention, then response times should not vary as a function of number of distractors in the feature condition. Further, if search for targets defined by a conjunction of features requires feature integration, and thus focal attention, then response times should vary as a function of number of distractors in the conjunction condition. Indeed this is what the researchers found. For the feature condition, response times remained relatively flat as the number of distractors increased, with search slopes of approximately 3 ms per item in the display. For the conjunction condition, response times increased linearly with the number of distractors, with search slopes of approximately 60 ms per item in the display. Treisman and Gelade interpreted these results to indicate that simple features can be processed in parallel across a visual scene, whereas focal attention is required to integrate multiple features.

According to feature integration theory, simple visual features are processed in parallel across a visual scene, whereas integrated representations that involve multiple features are constructed serially with the help of focal attention. Accordingly, the construction of visual representations can be seen as occurring in a hierarchical fashion, from the extraction of simple features, to the construction of whole objects, to the extraction of meaning. Treisman and Gelade (1980) demonstrated that the identity of a

feature can be extracted without awareness of its exact location, whereas the identity of a conjunction of features cannot be extracted unless its location is also known. To demonstrate this additional point, participants were presented with two rows of six objects each. The distractor objects consisted of the letters ‘O’ in pink and ‘X’ in blue. In the feature condition, the target object could either be an ‘H’ presented in pink or blue, or an ‘X’ or ‘O’ presented in orange. In the conjunction condition, the target object could be an ‘X’ presented in pink or an ‘O’ presented in blue. Notice, similar to the previous experiment, the targets in the conjunction condition were distinct only when considering the conjunction of features, whereas in the feature condition only one feature distinguishes targets from distractors. In this experiment, participants were asked to accurately localize and identify the target object.

If focal attention is required to process the conjunction of features, it should also be the case that successful identification of objects in the conjunction condition will depend on successfully localizing the target object. In addition, if simple features are processed in parallel, identification of objects based on a single feature should not require successful localization. Accordingly, the researchers found that in the conjunction condition, when participants failed to locate the target object correctly, they were also unable to identify that target correctly. In the feature condition, even when participants failed to localize the target correctly, they were above chance in identifying that target correctly. These results provide further evidence that simple features are processed in parallel outside the focus of attention. Conjunctions of features, on the other hand, require focal attention to create an integrated

object representation and ultimately to extract meaning. Presumably, the processing of more complex scenes operates in a similar fashion, from the parallel extraction of simple features to the serial extraction of meaning from whole objects.

Considerable work has also been done to examine the types of perceptual features that can capture attention, with a particular emphasis on the importance of singleton perceptual features; that is, perceptual features of objects that stand out because of their uniqueness relative to the perceptual features of neighbouring objects. Using the visual search paradigm, Theeuwes (1992) presented visual arrays containing 5, 7, or 9 objects. A line segment appeared inside each of the objects, and the task was to report the orientation of the line segment inside a target object that was defined by a particular singleton (or 'odd') feature.

There were two critical conditions: in the form condition, the target was the odd shaped object in the display; in the colour condition, the target was the odd coloured object in the display. Within each of these conditions, there were both distractor and no-distractor trials. For the form condition, distractor trials were those in which the odd shape was not the only singleton in the display; rather, there was also an odd coloured object in the display. As such, no-distractor trials might consist of a green circle target presented amongst green square distractors, whereas for distractor trials one of the distractors might be a red square rather than a green square. For the colour condition, distractor trials were those in which the odd coloured object was not the only singleton in the display; rather, there was also an odd shaped object in the display. As such, no-distractor trials might

consist of a green circle target presented amongst red circle distractors, whereas for distractor trials one of the distractors might be a red square rather than a red circle.

Theeuwes found that the response time search slopes for the distractor and no-distractor trials in both the form and colour condition did not differ from zero, indicating a parallel search occurred for both trial types. In other words, the singleton feature objects "popped out" of the field of distractors in all conditions. However, in the form condition, response times were slower for distractor than no-distractor trials across all display sizes. This result suggests that when searching for a unique form, a unique distractor colour captured attention and interfered with search for the unique form. In contrast, in the colour condition, response times for distractor and no-distractor trials did not differ, suggesting that unique shaped distractors did not capture attention when searching for an odd coloured target. Theeuwes interpreted these results to indicate that as feature information is extracted in parallel across a visual scene, unique singleton perceptual features can capture attention, but that whether or not this attention capture occurs can depend on the relative salience of the singleton features. In any particular context, some singleton features will be more salient than others, and attention capture will hinge on this relative salience. The time course of processing in the particular case described above was such that singleton colour objects were more salient than singleton shape objects.

In summary, based on these (Theeuwes, 1992) and other findings (Theeuwes, 1991; 1994a), Theeuwes (1994b) has suggested that as attention is allocated in parallel across a visual scene, which features capture focal attention is dependent on the relative salience of

those features (although see Yantis and Jonides, 1984 who suggest abrupt onset may be a special case). Note, attention capture in this sense is in accord with early selection models of attention in that, although the feature (e.g., colour or form) that captures attention may change from one visual scene to the next, it is the relative salience of perceptual features that drives attention capture.

Folk, Remington, and Johnston (1992), however, have argued that processes other than the relative perceptual salience of features may dictate attention capture. According to these researchers, task constraints can create an attentional set that biases the attention system to respond exogenously to some features, but not others. To demonstrate this point, Folk et al. used a modified spatial cueing task in which a target could appear in one of four spatial locations. On some trials an abrupt onset cue was presented in either the same (valid) or different (invalid) location as the target. In one condition, the target was also an abrupt onset and participants had to indicate its identity. In a second condition, the target was a different colour from a group of distractor items. Again, participants were to indicate its identity. In this way, the cue shared a common feature (abrupt onset) with the target in one condition, but not the other.

According to Folk et al. (1992), the constraints of the task (e.g., looking for an abrupt onset target) create an attentional set. If the cue contains a feature that matches the attentional set, then it will capture attention. In the different colour target condition, the attention system would be set for an odd coloured target, and therefore an abrupt onset cue should not capture attention. In contrast, in the abrupt onset target condition, the attention

system would be set for an abrupt onset target, and therefore an abrupt onset cue should capture attention. Indeed, only in the abrupt onset target condition were participants slower to respond to targets following invalid cues than to targets following no cue. This finding suggests that attention was drawn to the abrupt onset cue in an exogenous fashion only in the abrupt onset target condition. Moreover, in a second experiment, Folk et al. demonstrated the reverse effect when using an odd colour cue rather than an abrupt onset cue. These results indicate that exogenous attention can be modulated by endogenous factors. Specifically, properties of the task create an attentional set, and exogenous attention is then sensitive to features that match that attentional set.

The experiments reviewed above offer examples of how attention capture has been measured using visual search and spatial orienting tasks. The results from these experiments are largely consistent with tenets of early selection models of attention. Specifically, perceptual features of stimuli are assumed to be processed in parallel across a visual scene, and detection of features does not require focal attention. The experiments by Treisman and Gelade (1980) further confirm that feature conjunctions, the basis of complex object perception, cannot be processed in parallel, but rather require the focus of attention. An important implication of these tenets of the early selection model is that the extraction of meaning requires focal attention. Several studies that have examined this issue are reviewed in the following section.

Attention Capture by Semantic Features

One way to test whether semantic features can capture attention is with a modified visual search task. An early study that used this approach was reported by Brand (1971), who presented participants with a matrix of random letters (6 columns by 35 rows). Embedded within the matrix was a target that was either a pre-specified letter that differed from the distractor letters, a pre-specified number, or an unspecified number. Note, on some trials the target was semantically similar to the distractor items (i.e., both were letters), whereas on other trials the target was semantically dissimilar from the distractor items (i.e., the target was a number and the distractors were letters). Participants were tasked with scanning the matrix, from top to bottom, until they found the target. Brand argued that letters and numbers represent semantically different categories, but their perceptual features are similar. Therefore, if search times are faster when the target is a number amongst letters than when the target is a letter amongst letters, this result would indicate that attention is captured by semantic, not perceptual characteristics. Indeed, this is the result that was observed. Some have argued, however, that letters and numbers are not entirely perceptually comparable (Jonides & Gleitman, 1972). Therefore, the faster search times for a number amongst letters, than for a letter amongst letters, might reflect attention capture by perceptual feature differences.

In an attempt to focus only on semantic differences between these stimulus classes, Jonides and Gleitman (1972) introduced a target ‘O’, presented amongst a group of distractors that were made up of either letters or numbers. Each visual display could

contain 2, 4, or 6 items. Importantly, prior to the presentation of the search display, the target 'O' was specified as being either the letter "O" or the number "zero". Therefore, the perceptual features of the target were held constant on all trials, but the meaning of the target varied across trials. The crucial result was that there was a flat response time search slope when the semantic category of the target differed from that of the distractors, but a linearly increasing slope when the semantic category of the target matched that of the distractors. This result demonstrates that the target captured attention only when its semantic category was unique amongst the distractors, despite controlling for perceptual feature similarity between the two search context conditions.

Although the Jonides and Gleitman (1972) result was an important one in the field, it has since been shown to be difficult to replicate (Duncan, 1983). As such, an answer to whether attention can be captured by semantic properties of a visual scene requires converging support from other methods. Using a clever methodology, Smilek, Dixon, and Merikle (2006) attempted to avoid the issues raised in Brand's (1971) experiment while still maintaining semantic differences between target and distractor items. To do so, the researchers trained participants to associate a first one-word label (e.g., pencil) with two line segment orientations: a target line orientation (e.g., left) and a distractor line orientation (e.g., horizontal). They also trained participants to associate a second one-word label (e.g., elephant) with two additional line segment orientations, one a target (e.g., right) and the other a distractor (e.g., vertical). Each search display included only one of the two target types and one of the two distractor types. In this way, the target could be physically

different, but categorically the same as distractor items, or the target could be both physically and categorically different than the distractor items. The clear prediction is that if meaning 'oddness' can capture attention, search times should be more efficient for targets that are both physically and categorically different from the distractor items than for targets that differ only in perceptual features. Indeed, search slopes that related response time to number of distractors were significantly shallower for targets that were both physically and categorically different from distractor items than for targets that were different physically but categorically the same as the distractor items. These results suggest that search was more efficient for objects that differed in meaning from the surrounding items and appears to be a violation of an important assumption posited under early selection models of attention, namely that attention is required to extract meaning.

The research reviewed to this point suggests that salient perceptual features of stimuli can capture attention (Theuwes, 1991; 1994b; Treisman & Gelade, 1980). Further, the physical features that capture attention may be affected by the attentional set of the observer, which implies that the viewer's goals can modulate attention capture from one task to the next (Folk et al., 1992). These observations are in accord with Broadbent's (1958) early selection model of attention in that they suggest that only the low level perceptual features of objects are processed pre-attentively. In contrast to early selection models of attention, some researchers have attempted to show that attention can be captured by semantic properties of a visual scene (Brand, 1971; Jonides & Gleitman, 1972). These studies have been subject to criticism for stimulus set confounds, specifically

for not controlling the physical properties of the stimuli while manipulating their semantic properties (Duncan, 1983; Smilek et al., 2006). Aside from the study of Smilek et al. (2006), there is scant evidence in the basic visual search literature that semantic properties guide the capture of attention (although see Lachter, Ruthruff, Lien, & McCann, 2008 for a demonstration using the Stroop paradigm). The majority of these studies use visual arrays composed of relatively simple objects defined by one or more visual features (e.g., colour, shape, abrupt onset). There are, however, a number of studies using more complex stimuli to investigate scene perception that may shed further light on the extent to which attention is attracted by object features.

Complex Scene Perception

As mentioned earlier, studies in inattentional blindness and change blindness suggest that attention is selective and that representations of complex scenes are created in a piecemeal fashion from one saccade to the next. The question remains, however, which areas of a scene are prioritized when viewing a complex scene? Moreover, is this selectivity driven exclusively by the physical features of the stimuli, as suggested by the studies using simple visual arrays? To this point, the evidence for semantic attention capture using complex scenes has been varied. There are two main areas of research in scene perception that may shed some light on these questions. The first involves investigating how the viewer recruits the information in complex scenes. According to the early selection model of attention, simple physical features of the scene might be expected to be extracted first, before features are integrated into objects in a hierarchical fashion

until the scene representation is complete. Research aimed at investigating this question has focused on the time course of information extraction when viewing complex scenes. The second involves investigating what areas of a complex scene preferentially guide or capture attention.

Representing Scene Category Information and Object Information

There is general agreement that when viewing complex visual scenes, the category information of the scene is recruited rapidly, before the processing of local object information occurs. Moreover, the category information is represented separately from object information (Sampanes, Tseng, & Bridgeman, 2008). Although early research has placed the recruitment of scene category level information as occurring at between 100-120 ms following scene onset (Biederman, 1981; Intraub, 1980; 1981; Metzger & Antes, 1983; Potter, 1975; 1976; Potter & Levy, 1969), more recent research indicates that the recruitment of category information can occur much earlier (Castelhano & Henderson, 2008; Oliva & Schyns, 1997; Schyns & Oliva, 1994; VanRullen & Thorpe, 2001). For example, Castelhano and Henderson briefly presented a photograph of a natural scene for an exposure time of 25, 33, 42, 50, or 250 ms. The image was followed by a visual mask, created by scrambling another scene image. Following the presentation of the scene and mask, a word was shown describing an object that was either semantically congruent or incongruent with the preceding scene. Participants were required to indicate whether an object matching the word was present or absent in the preceding scene. Although there was never a target object (congruent or incongruent) matching the written description in the

preceding scene, the researchers used the present/absent response as a way to calculate response bias. A response bias towards target presence was taken as an indication that the category of the scene had been processed. Interestingly, a positive response bias in favour of object presence for congruent objects emerged at a scene presentation time of 42 ms. This result was taken as an indication that viewers can process the category information of a scene very early into scene processing.

On the other hand, the processing of local features, such as the individual objects that make up a scene, has been shown to occur slightly later (Fabre-Thorpe, Delorme, Marlot, & Thorpe, 2001; VanRullen & Thorpe, 2001). For example, using a go/no-go task, VanRullen and Thorpe presented participants with images containing target animals (go trials) and images containing distractor vehicles (no-go trials) in some blocks of trials and the opposite target-distractor configuration in other trials, alternating from one block to the next. The alternating configuration of target-distractor blocks allowed the researchers to compare performance on go trials with no-go trials within a target category, but also across target categories. Measuring event-related potentials (ERP), the researchers found differential activity in frontal electrodes at about 75 ms when combining target/non-target images of the same category. In line with the literature reviewed above, this finding implies that scene categorization occurs rapidly. However, when comparing the processing of target images with non-target images across categories, differential activity in frontal areas was found at about 150 ms. This finding suggests that object processing begins slightly later

than scene categorization. Interestingly, the time course of object processing has been shown to be immune to training, even after 3 weeks (Fabre-Thorpe et al., 2001).[1]

The rapid processing of a scene's category, prior to the processing of local items, may at first blush appear to be a violation of early selection models of attention. Indeed, beyond the processing of a scene's general category, the structure inherent in a particular context naturally constrains certain areas of the scene to be of higher interest than others (e.g., a street in a landscape scene). Rensink et al. (1997) have argued that a viewer's past experience coupled with the structure of a scene context guides attention to areas of high interest. Rensink et al. demonstrated this point by presenting images using a change detection task, similar to the change blindness task described earlier, wherein across successive presentations of an image an object is removed and then replaced. Rensink et al. showed that participants were more efficient at detecting changes in areas of the scene that had been rated as high interest by an independent group of participants. This result, in conjunction with the results supporting rapid scene categorization, may be taken as an indication that scene category information is processed automatically and areas of high interest attract attention. This assertion constitutes a clear violation of early selection models of attention, which assume that semantic level information is a product of focal attention, and therefore cannot attract attention in the first place.

It should be noted, however, that many researchers have asserted that rapid recruitment of scene category information is derived from one or more local features of the scene (De Graef, Christiaens, & d'Ydewalle, 1990; Fei-Fei, Iyer, Koch, & Perona, 2007;

Friedman, 1979; VanRullen & Thorpe, 2001). That is, processing a few local features of the scene may summon schemas of similar scenes. Thus, rapid processing of scene category information is not necessarily a violation of early selection models of attention. Take as example an experiment reported by De Graef et al. (1990). Participants were asked to explore a series of complex line drawings for the presence of non-existent objects. Each image contained two targets, with each target fitting into one of five scene context violation conditions. The target object could represent a non-violation (i.e., semantically congruent with the scene), a support violation (i.e., placed in a physically implausible location), a size violation (i.e., unnaturally small or large given its location in the scene), a position violation (i.e., placed in a physically plausible, but contextually inappropriate location), or a probability violation (i.e., semantically incongruent with the scene context). The time course of target fixations revealed longer gaze durations for targets that violated probability, position, and support when these objects were first fixated late into scene processing than when they were fixated early during scene processing. The researchers interpreted these findings to suggest that the construction of scene representations occurs in stages, beginning with coarse object characteristics (e.g., size) being processed in parallel, early during scene processing and prior to the processing of more detailed features.

Benefits of Context Congruency on Object Processing

Another way to demonstrate attention capture by semantic information would be to show preferential processing for objects that are semantically incongruent with the scene context. However, much of the evidence in scene perception has shown performance

benefits of object/context congruency. For example, Neider and Zelinsky (2006) presented participants with a series of computer generated complex images depicting a variety of desert scenes, with a mountain range separating the top of the image from the bottom. Each scene contained six objects, made up of three object types: a jeep, a blimp, and a helicopter. The jeep objects always appeared on the bottom half of the screen (i.e., the ground), the blimp objects always appeared on the top half of the screen (i.e., the sky), and the helicopter objects appeared in either top or bottom half of the screen. At the beginning of each trial, participants were given a written description of a target object (e.g., BLUE JEEP) and they were to determine whether a target object matching that description was embedded in the subsequent scene. Neider and Zelinsky reported that participants were significantly slower at detecting the presence of the contextually ambiguous object (i.e., helicopter) than the contextually constrained objects, but that there was no difference in search times on target absent trials. Moreover, the participants' first eye saccade was more likely to be directed towards contextually appropriate areas of the scene for the contextually constrained objects, regardless of whether or not the target was present. The researchers interpreted these results to indicate that attention was guided by top-down influences (e.g., past experience with similar contexts), not the physical features (i.e., the presence) of the objects.

Other studies have shown that the context of a scene can help participants identify congruent objects. Palmer (1975) presented participants with a complex line drawing of a particular scene context for 2 s, followed by a 1300 ms inter-stimulus interval (ISI), prior to

the presentation of a line drawing of an isolated object that was either congruent or incongruent with the preceding scene, for 20, 40, 60, or 120 ms. Participants were then given 20 s to write down the type of object that was just presented. Even at the briefest exposure times, participants were more accurate at identifying the isolated object when it followed a congruent context than an incongruent context. This result suggests that the context of a scene prepares a viewer for objects that are likely to be encountered in such a scene. In this way, a congruent context creates expectations regarding what types of objects should be present.

Moreover, viewers have been shown to be able to quickly categorize objects by using scene information (Sun, Simon-Dack, Gordon, & Teder, 2011). Using a go/no-go task, Sun et al. presented participants with a 20 ms exposure of an object that was either an animal or a vehicle. The objects were presented either in isolation, embedded within a congruent context, or embedded within a phase-randomized scene. Participants were asked to make a response on trials in which an animal was present (go trials) and withhold a response on trials in which a vehicle was present. In this case, participants were significantly faster at responding on go trials if the object had been embedded within a congruent context than presented in isolation or in a phase-randomized scene. This finding further suggests that the category information of a scene is processed early during scene processing and that this information can be used for efficient processing of congruent objects.

Benefits of Context Incongruency on Object Processing

The studies reviewed above suggest that a congruent context can help guide attention to important areas of a scene (Neider & Zelinsky, 2006), and that a congruent context can help with object identification (Palmer, 1975) and object categorization (Sun et al, 2011). However, there are a number of studies reporting performance benefits for objects embedded in an incongruent context. In a seminal study on scene perception, Loftus and Mackworth (1978) presented participants with a series of complex line drawings depicting a variety of scene contexts. In order to create both semantically congruent and incongruent scenes, the researchers took a pair of scenes and swapped an object in one for an object in another. For example, one scene might depict a farmyard with a tractor, whereas another scene might depict an underwater context with an octopus. Each of these scenes would represent congruent contexts. To create a pair of incongruent contexts, the researchers placed the tractor in the underwater scene and the octopus in the farmyard scene, thereby matching the target objects across the context condition. Participants were told to use the 4 s presentation time to freely explore the scene in preparation for a later scene recognition memory test. Under these conditions, Loftus and Mackworth found that the semantically incongruent objects were fixated earlier than the semantically congruent objects. Moreover, the eye saccades towards the incongruent objects were significantly larger than those directed towards the congruent objects. The researchers took these findings as an indication that semantically incongruent objects are

processed peripherally and attention is attracted to these objects early into scene processing.

The findings and interpretation offered by Loftus and Mackworth (1978) have been considered controversial, in part because they suggest scenes are processed in parallel and that semantically incongruent objects attract attention early into scene processing. This assertion would seem to violate basic tenets of early selection models of attention. Moreover, the findings reported by Loftus and Mackworth have been difficult to replicate (D'Graef, Christiaens, & D'Ydewalle, 1990; Friedman, 1979; Gareze & Findlay, 2007; Henderson, Weeks, & Hollingworth, 1999; Rayner, Castelhano, & Yang, 2009; Vo & Henderson, 2009; 2011). For example, using a similar task, Henderson et al. (1999) presented participants with a series of complex line drawings, in this case for 15 s each, but also under threat of a later scene recognition memory test. Just as was the case in Loftus and Mackworth's experiment, each image contained a target object that was either semantically congruent or incongruent with the context of the scene. Despite the similarities in task, Henderson et al. (1999) report participant's eye gaze was equally likely to fixate the congruent targets as they were the incongruent targets. Furthermore, the size of the saccades directed towards the target object was larger for congruent targets than incongruent targets. These findings cast doubt on the both the findings and interpretation reported by Loftus and Mackworth. Specifically, the results of Henderson et al. (1999) do not support the notion that attention is allocated preferentially to semantically incongruent objects.

The findings reported by Henderson et al. (1999) suggest that attention may be guided towards objects that fit the context of a visual scene. This idea is also supported by a large group of studies showing that a viewer uses the rapid recruitment of a scene's categorical information to guide attention to areas that are relevant to their search. For example, Vo and Henderson (2011) gave participants a short (250 ms) preview of a computer generated complex scene containing a target object that was either semantically congruent or incongruent, followed by a pattern mask for 50 ms. Next, participants were given a one word description of a target object. Finally, the scene was presented again, but viewing was restricted to a 5-degree diameter window that moved in conjunction with the participants' eye movements. Participants were tasked with finding an object that matched the written description as quickly and accurately as possible. Although Vo and Henderson found no difference in the latency or size of the first saccade to enter the target area, participants were significantly faster at finding congruent targets than incongruent targets. Vo and Henderson interpreted these results to indicate that the brief preview of the scene allowed participants to take advantage of the scene's contextual structure to efficiently locate objects that were congruent with the scene.

In contrast to the results reported by Loftus and Mackworth (1978), those reported by both Henderson et al. (1999) and Vo and Henderson (2011) offer no evidence that semantically incongruent objects capture attention. Recall, however, that the results reported by Loftus and Mackworth also appear to indicate that semantically incongruent objects capture attention early into scene processing. In an effort to investigate specifically

the processing of objects early during scene viewing and any influence scene context may exert, Gordon (2004) presented a series of complex line drawings for either 53 or 147 ms. Each of the images contained both a congruent and incongruent target object. The scenes were immediately followed by a mask containing a probe that was either at the same location of the congruent target, at the same location of the incongruent target, or at a location where there was no object in the preceding scene. Gordon reasoned that if attention was attracted to semantically incongruent objects early during scene processing, identification of the probe should be better for those probes appearing in the same location as the preceding incongruent object. Indeed, this is what he found - participants were more accurate at identifying the probe when it was located in the same area as the incongruent object than when it was located in the same area as the congruent object in the preceding scene. This difference, however, was present when the scene image was presented for 147 ms, and not present for scenes presented for 53 ms. From these results, Gordon argued that within 147 ms of scene processing, attention is captured by semantically incongruent objects.

Benefits of Semantic Incongruency in Change Detection

Several studies using a change detection task have also reported results that are consistent with the idea that semantically incongruent objects attract attention. A particular variant of the change detection task, first developed by Rensink et al. (1997), has been useful in the investigation of this attention attraction hypothesis. In this task, a scene containing a target object is presented briefly (e.g., 250 ms), followed by the presentation

of the same scene without the target object for an equally brief exposure. An important element of this task is that an ISI screen is presented between the scene presentations. The ISI screens are crucial in disrupting an extended exploration of the scene and have been argued to mimic elements of natural scene viewing conditions, such as eye blinks and saccade shifts. Using this task, studies have shown that some areas of a scene are prioritized over others, such as areas that are independently rated as interesting (Rensink et al., 1997), relevant based on prior experience (Werner & Thies, 2000; Jones, Jones, Smith, & Copley, 2003), or informative (Hollingworth & Henderson, 2000).

Hollingworth and Henderson (2000) used the change detection task to investigate the influence of context congruency on scene viewing behaviour by presenting participants with a series of complex line drawings containing target objects that were either semantically congruent or incongruent with the scene context. The images were presented for 250 ms, and separated by blank screens of 80 ms each. In one condition, the change from one scene presentation to the next was the deletion of the target object. In another condition, the change from one scene presentation to the next was the orientation of the target object. In both of these conditions, the target object could either be semantically congruent or incongruent with the scene context. In a third condition, there was no difference across the image presentations. The images from the three change conditions and the two context conditions were presented in a random order. Participants were tasked with indicating whether a change was present from one scene presentation to the next as quickly as possible. Across change conditions (i.e., deletion and orientation), participants were

significantly faster at detecting changes to incongruent than congruent objects, although there was no difference in terms of response accuracy across the context conditions. These results were interpreted to suggest that objects that are semantically incongruent with the scene context are particularly informative because they share little-to-no overlapping information with other objects in the scene. Congruent objects, on the other hand, are less informative because they share overlapping information with other objects in the scene. Moreover, the informative nature of the incongruent objects gives them an attentional priority.

Hollingworth and Henderson (2000) discussed at length two accounts for why incongruent objects might have attentional priority over congruent objects in a change detection task. First, it may be that semantic disfluency of the incongruent target attracts attention, thereby producing faster response times for those objects (although see Gordon, 2006). The researchers termed this interpretation the *attention attraction hypothesis*. According to this view, upon recruiting the categorical information of the scene, the semantically incongruent object causes a disfluency in the scene representation, thereby attracting attention for further investigation. Second, rather than attention being attracted to semantically incongruent objects, it may be that attention fails to disengage from incongruent objects. The researchers termed this interpretation the *attention disengagement hypothesis*. According to this view, attention is allocated randomly and serially about the scene, however, once attention lands on an incongruent object it has a tendency to linger, perhaps in an attempt to reconcile the identity of the object with the contextual information

of the scene. As attention lingers on the semantically incongruent object, the next scene presentation is likely to occur, making it more likely that the change will be detected.

It is important to note that the attention disengagement hypothesis detailed by Hollingworth and Henderson (2000) is perfectly in accord with early selection models of attention, which assume that attention cannot be captured by semantic characteristics, but rather that attention is required to extract meaning from an object. The attention disengagement hypothesis suggests a similar view of scene perception in that attention is not attracted preferentially to semantic incongruency, but rather lingers longer on these objects, perhaps as meaning is extracted. In contrast, the attention attraction hypothesis would seem at odds with early selection models of attention, as it suggests that attention is captured by a disfluency between the semantic characteristics of an object and the semantic characteristics of the scene context. The attention attraction hypothesis assumes that some form of discontinuity in meaning across the visual scene is precisely what attracts attention.

Summary

A number of studies have demonstrated that the salience of perceptual features can capture attention in an automatic and effortless fashion. Many of these studies have used visual arrays composed of simple objects with just one or a small number of basic perceptual features. Investigations of attention capture by semantic characteristics using such methods have produced results that are less than clear, with some studies showing category level effects, but others not. The research in complex scene perception has revealed equally mixed results. An early study by Loftus and Mackworth (1978) revealed

that semantically incongruent objects with respect to the context of the scene in which they were embedded captured attention early during scene processing. This finding, however, has been difficult to replicate. Studies dedicated to isolating the early moments of scene perception have also shown some indication that semantically incongruent objects capture attention early during scene processing. The object characteristics that capture attention in these situations, however, remain an issue of debate. Finally, a number of studies using a change detection task, which employs brief and interrupted scene presentations, have shown more efficient behaviour for semantically incongruent objects than congruent objects. Competing hypotheses describing the processes that underlie performance benefits for incongruent objects have been offered, leaving the issue in need of further study.

Overview of the Empirical Chapters

Given the conflicting results in studies investigating scene perception, with some showing congruency benefits and others showing incongruency benefits, both within tasks (Loftus & Mackworth, 1978; Henderson et al., 1999) and across tasks (Vo & Henderson, 2011; Gordon, 2004; 2006), there is a need to examine whether multiple processes are at work when viewing complex visual scenes. In particular, it may be that some scene perception processes benefit from a congruent context, perhaps to guide attention to important areas of the scene. However, other processes may benefit from an incongruent context, perhaps to help with detection of novelty. As a way to investigate whether multiple processes underlie behavior when viewing complex scenes, in the second chapter I manipulate the parameters of a change detection task, a task that has typically produced

incongruency benefits, in a way to emphasize processes that support object detection (Experiment 1) and processes that support object identification (Experiment 2). Using this approach, I demonstrate that congruency benefits can also be measured using a change detection task, when the processing emphasis is shifted away from target detection and toward target identification.

As noted, the conventional change detection task typically produces incongruency benefits. Two competing hypotheses have been suggested to account for these costs - attention attraction and attention disengagement (Hollingworth & Henderson, 2000). The attention attraction hypothesis suggests that attention is allocated in parallel across the scene and that semantically incongruent objects attract focal attention. In contrast, the attention disengagement hypothesis suggests that attention is allocated in a serial fashion across the visual scene, but once attention has landed on a semantically incongruent object, it has a tendency to linger on that object for further processing. In Chapter 3, I use a change detection task to demonstrate incongruency benefits, while monitoring eye movements to assess the nature of attentional processing. Using a number of eye movement measurements, I demonstrate that attention does not linger on semantically incongruent objects, but rather that attention is attracted to these objects.

Also as noted, some studies assessing scene perception have demonstrated congruency benefits in the processing of target objects, whereas other studies have demonstrated incongruency benefits. These conflicting results suggest that multiple processes underlie behaviour when viewing complex scenes. Moreover, these conflicting

results suggest that the nature of the task may play an important role in dictating which scene perception processes predominate in any given study. For example, in many of the tasks used to demonstrate congruency benefits information about the target object is given prior to the presentation of the scene (e.g., visual search). In contrast, in many of the tasks used to demonstrate incongruency benefits no information is given about the target object prior to the presentation of the scene (e.g., change detection). To this point, there are no studies that carefully compare behaviour between a task that produces a congruency benefit and a task that produces an incongruency benefits. In Chapter 4, I introduce a component of a conventional visual search task to a conventional change detection task in an effort to shift the process weighting in favour of congruency benefits. In addition, using a controlled stimulus set, I compare performance in a conventional visual search task to performance in a hybrid task that has element of both change detection and visual search. Using this approach, I demonstrate congruency benefits when viewing a scene for object search and incongruency benefits when viewing a scene for object change.

A Note on Terminology

Throughout each of the following chapters I use the term *congruency benefit* to describe more efficient performance on trials in which the target object and scene context are semantically congruent compared with trials in which the target object and scene context are semantically incongruent. I use the term *incongruency benefit* to describe the opposite pattern of results; namely, more efficient performance on trials in which the target object and scene context are semantically incongruent compared with trials in which the

target object and scene context and semantically congruent. It is important to note that in none of the experiments described in the following chapters was there a neutral or baseline condition in which performance in the context congruency conditions were compared. As such, the term *benefit* refers only in contrast to the opposing context condition.

Footnote

Note 1. The timeline for scene and object processing noted here and in subsequent chapters refers to brain processing time, not perceptual processing or viewing time.

Chapter 2: Context Congruency Effects in Change Detection: Opposing Effects on Detection and Identification

LaPointe, M. R. P., Lupiáñez, J., & Milliken, B. (2013).

Visual Cognition, 21, 99-122.

Preface

The purpose of Chapter 2 was to identify the processes that underlie complex scene perception and to attempt to systematically weight these processes differentially within a task to produce opposing behavioural performance. This chapter contains three experiments. In Experiment 1, the robust finding of performance benefits for incongruent objects in a change detection task (Rensink, O'Regan, & Clark, 1997) was successfully replicated using a stimulus set composed of natural scenes. In Experiment 2, a change detection task was manipulated by removing the typically presented interleaving white screens, leaving the detection of changes trivial and based on transient luminance changes alone. Under these task parameters, we report performance benefits for semantically congruent objects. From these results, we reasoned that when change detection is trivial, performance in this task relies mainly on accurate object identification. Moreover, under these parameters, a congruent context should aid successful performance. The rationale for

Experiment 3 was to minimize the contribution of the object identification process by having participants localize, rather than identify the changing object. Under these parameters, we report performance benefits for incongruent objects, whether or not the interleaving white screens are presented between the scene images. The results reported in this chapter represent an important contribution to the literature on complex scene perception. Notably, these results constitute the first demonstration of both congruency benefits and incongruency benefits within a change detection task. Moreover, from these results we speculate that two independent processes underlie the perception of complex scenes.

Abstract

In tasks of visual change detection, researchers have shown that objects embedded in a contextually incongruent scene tend to be detected faster than objects embedded in a contextually congruent scene. This finding is curious given that a contextually congruent scene contains a host of cues that should aid in object perception, such as which types of objects are to be expected and their probable location. We replicated this context incongruency benefit (Experiment 1), but also showed that this effect reversed to a context congruency benefit when change detection was made trivial by reducing the inter-stimulus interval (ISI) between image presentations from a more conventional 250 ms to 0 ms (Experiment 2). Across Experiments 2 and 3, we also showed that context incongruency impeded change detection performance when the task required identification of the changing object, but aided change detection performance when the task required only localization of the changing object. These results highlight the contribution of two processes to change detection performance that are affected in opposite ways by context congruency. Whereas incongruent contexts appear to facilitate the process of detecting and localizing the object that is changing, congruent contexts appear to facilitate identification of the object that is changing.

Introduction

Processes that support the detection and identification of visual events are important in a wide range of everyday activities, from finding an empty seat in a lecture hall to crossing a busy street. How do we perform such complex behaviors with apparent ease? There is general agreement among researchers that a combination of bottom-up and top-down processes are at play when a viewer navigates a visual scene (Bacon & Egeth, 1994; Itti & Koch, 2000; Zoest, Donk, & Theeuwes, 2004). Bottom-up processes are those derived entirely from the perceptual properties of the visual scene itself, such as the visual salience of the object (Theeuwes, 1991; Yantis & Jonides, 1990), whereas top-down processes take advantage of things such as the viewer's prior experience with a scene's context (Bacon & Egeth, 1994; Chun, 2003; Green & Hummel, 2006). Taking the search for an empty seat example, the top-down goal of finding a seat directs the viewer's attention voluntarily. At the same time, the environment also provides bottom-up input, indicating perhaps where an oddball empty seat might be. Despite recognition that both top-down and bottom-up processes guide how we search visual scenes, the interaction between these classes of processes remains an issue of considerable debate (Itti & Koch, 2000).

Still at issue is how these processes are prioritized as a viewer first constructs a visual representation and compares that representation across gaze saccades and scene changes. Built upon the idea that top-down influences can bias scene recognition, Oliva and Torralba (2007) argue that the way in which we categorize scenes at a glance is directly

mediated by our understanding of the context of that scene. That is, the context gives the viewer important information pertaining to what types of objects are expected, as well as their location and interaction with other objects in the scene (Chun, 2003; Gordon, 2004; Green & Hummel, 2006; Oliva & Torralba, 2007). By this view, scene categorization takes place quickly, with viewers able to grasp semantic information associated with the general category of a scene, or the scene's gist, within the first 100 ms of viewing (Biederman, 1981; Fei-Fei, Iyer, Koch & Perona, 2007; Henderson & Hollingworth, 1999; Intraub, 1981; Potter, 1975; 1976; Oliva & Schyns, 1997; Sampanes, Tseng & Bridgeman, 2008; Schyns & Oliva, 1994). Indeed, gist recruitment has been reported to require as little as 42 ms (Castelhano & Henderson, 2008), while recognition of individual objects that make up a scene has been reported to occur as quickly as 150 ms (VanRullen & Thorpe, 2001). These results fit with the general view that rapid recognition of the gist of a scene can guide subsequent scene exploration (Castelhano & Henderson, 2008; Sampanes et al., 2008).

Benefits of Context Congruency

Using a contextual cueing procedure, Chun and Jiang (1998) showed that viewers can learn the global configuration of a visual scene and then translate that learning into faster search responses for targets embedded in the same displays compared with new displays. That is, learning about the context in which the targets were previously encountered influenced the way in which viewers searched for targets in those same contexts on subsequent trials. A role for prior experience in scene perception has also been

shown using natural scenes. Testing football players, Werner and Thies (2000) showed that these viewers were less susceptible than non-football players to change blindness for targets appearing and disappearing from football scenes (for other demonstrations of expertise effects in change detection see Curran, Gibson, Horne, Young, & Bozell, 2009; Jones, Jones, Smith, & Copley, 2003). In similar fashion, Rensink, O'Regan, and Clark (1997) demonstrated how the global properties of a scene can constrain semantic points of interest. In this experiment, participants detected changes relatively efficiently in areas of the scene that had been deemed to be of high interest by an independent group of raters. Taken in combination, these results show the strong influence of prior experience, expertise, and scene schematics on the deployment of focused attention.

There is also ample evidence from the visual search literature that the context in which targets are embedded affects viewers' ability to find and identify those targets (Davenport & Potter, 2004; Henderson, Weeks, & Hollingworth, 1999; Kelley, Chun, & Chua, 2003; Neider & Zelinsky, 2006; Palmer, 1975; Sun, Simon-Dack, Gordon, & Teder, 2011; Vo & Schneider, 2010). For example, Palmer (1975) showed that categorization of a target object is more efficient following presentation of a contextually congruent scene, and less efficient following presentation of a contextually incongruent scene, than following presentation of a blank scene (see also Sun et al., 2011). Similarly, Henderson and colleagues (1999) gave participants an object label and then asked them to find that target object in a visual scene that followed. Target objects that were contextually congruent with

the background scene were found faster and fixated earlier than target objects that were contextually incongruent with the background scene.

Benefits of Context Incongruency

The results summarized above show clearly that the context of a scene, in combination with prior knowledge about such scenes, can help identification of the object by constraining its categorization. Context can also help in finding an object by constraining attentional processes, and thereby produce performance benefits for target objects that are congruent with the scene context. Yet, there is a growing set of studies in the change detection literature that report the opposite effect (Brockmole & Henderson, 2008; Gordon, 2004; 2006; Hollingworth & Henderson, 2007; 2000; Hollingworth, Williams, & Henderson, 2001; LaPointe, 2011; Pezdek, Whetstone, Reynolds, Askari, & Dougherty, 1989). For example, using the flicker paradigm (Rensink et al., 1997), Hollingworth and Henderson (2000) presented complex line drawings of natural scenes with one object changing across consecutive views of the images. The task was to detect the changing target object, and the target objects were either congruent or incongruent with the background image. The authors found that performance was faster for incongruent than congruent objects.

Why would change detection tasks lead to performance benefits for context incongruent targets while other search tasks produce the opposite pattern of performance? One possibility is that attention capture processes are especially important to detect and localize the changing object, and this attention capture process is particularly sensitive to

inconsistency between an object and the scene in which it is embedded. For example, it seems possible that attention is attracted in change detection tasks to areas in a scene in which rapid gist-based processing detects a semantic conflict between the scene context and an incongruent target embedded within that scene. In contrast, in tasks that do not require shifts of attention to detect changing targets, and that instead load heavily on object identification processes, it seems reasonable that rapid gist-based processing would generate a perceptual hypothesis about the possible object to be found, which would in turn benefit identification of context congruent objects.

The notion that some form of attention capture is critical to the benefit for incongruent objects found in change detection tasks finds support from the work of Hollingworth and Henderson (2000). They described their contextually incongruent targets as being semantically informative points in the scene, whereas the contextually congruent targets were considered to be semantically uninformative. That is, the incongruent objects did not interact with the other objects making up the scene in a natural way, creating a point of semantic importance. In line with this general idea, other semantic categories have also been shown to capture attention in change detection tasks, such as people (Bracco & Chiorri, 2009) and faces (Weaver & Lauwereyns, 2011). If areas of interest in a scene are attended to preferentially over areas of less interest, than changes that occur in those areas of interest ought to be detected with greater ease than changes in other areas of a scene (Hollingworth, Schrock, & Henderson, 2001; Rensink et al., 1997).

The Present Study

Although other researchers have noted that a form of attention capture may contribute to the context incongruency benefit often observed in change detection tasks, it remains unclear why this attention capture process predominates in change detection tasks but not in other tasks. There are two broad answers to this question that are of direct interest to us in this study. One possibility is that the requirement to detect change necessarily engages the gist-based attention capture process responsible for context incongruency benefits, rather than other context-sensitive processes that facilitate performance for context congruent targets. By this general view, one would expect that all forms of tasks that require change detection ought to produce a benefit for context incongruent over context congruent targets. An alternative possibility is that it is not the requirement to detect change per se that is critical, but rather that the typical change detection task magnifies the contribution to performance of the gist-based attention capture process to a point that it predominates over other processes that might well produce context congruency benefits.

In the spatial cueing literature in which the only context is a spatial cue that is presented just before the target, it has been shown that the spatial context (i.e., the cue) can at the same time hinder target detection (leading to faster detection responses to targets appearing at a location away from the spatial cue), but facilitate target discrimination (leading to faster identification responses to targets appearing at the same location as the

cue; Lupiáñez, Ruz, Funes & Milliken, 2007). We expected similar results for the semantic context manipulation in the current set of experiments.

In particular, rather than manipulating task (detection vs. discrimination), we measured context congruency effects across several variants of a change detection task. Across these change detection tasks, task parameters were altered that ought to shift the relative contributions of detection and identification processes to task performance in predictable ways. We were particularly interested in whether it would be possible to selectively alter the relative contributions of two processes: (1) an attention capture process that we presume hinges on gist-based processing, and that is facilitated by target-context incongruency; and (2) a target identification process that we presume is facilitated by target-context congruency.

Experiment 1

The goal of the first experiment was to establish a change detection procedure that reliably produces a context incongruency benefit (Brockmole & Henderson, 2008; Hollingworth & Henderson, 2000; Loftus & Mackworth, 1978). We used stimuli that had been used in a prior study by LaPointe (2011). In that study, a context incongruency benefit was indeed observed, with context congruency manipulated across separate blocks of trials. Here, context congruency was mixed randomly across trials rather than blocked. Given results of the prior study using these stimuli (LaPointe, 2011), results of prior work on rapid gist-based scene categorization (Biederman, 1981; Henderson & Hollingworth, 1999; Intraub, 1981; Potter, 1975; 1976; Oliva & Schyns, 1997; Schyns & Oliva, 1994),

and results from studies on the influence of semantic informativeness on attentional allocation (Hollingworth & Henderson, 2000), we predicted that changing objects embedded in an incongruent background would be detected faster than changing objects embedded in a congruent background. Furthermore, we predicted that participants would fail to detect change objects embedded in an incongruent background less often than changing objects embedded in a congruent background.

Method

Participants. Twenty undergraduate psychology students from McMaster University volunteered to participate in exchange for partial course credit.

Stimuli. The images used for all three experiments in the current study were used in a prior study by LaPointe (2011), and were created from photographs taken in Brisbane, Australia. For both within-subject conditions, 70 pairs of images were presented. Each pair included a background image (*A'*) and a background-plus-target image (*A*). The background images were created from photographs of natural scenes taken at several times of day and featured a variety of lighting conditions. The photographs varied widely in their level of complexity, including multiple landscapes, both indoors and outdoors. The background-plus-target images were created by digitally superimposing a target object onto a copy of one of the background-only images. The target objects included both animate and inanimate objects; featuring both human and non-human animals, plants, fixed objects (e.g., a lamp post), and moveable objects (e.g., a car).

The backgrounds on which the target objects were superimposed were either contextually congruent or contextually incongruent with the target objects. For the contextually congruent object/background pairing, a background was chosen such that the target object could be placed in a naturally fitting location (e.g., a toaster on a kitchen counter). For the contextually incongruent object/background pairing, a background was chosen such that the target object was placed in a scene where it would not typically be found, but in a naturally fitting location within that scene (e.g., an emu sitting on a living room floor). What was considered contextually congruent and incongruent necessarily differed across target object categories. While a forest scene may be contextually appropriate for a bear, it would be considered contextually inappropriate for a boat. It is important to note that although different backgrounds were chosen for a particular target object in the congruent and incongruent context conditions, the target object was always placed in a similar location in these two scene contexts. For example, if a bear was placed in the bottom right-hand corner of an image in the congruent context condition (e.g., a forest floor), it was also placed in the bottom right-hand corner of an image in the incongruent context condition (e.g., a bedroom floor).

Materials and Procedure. The stimuli were presented on a 17" Apple eMac using Livecode programming software. Each participant sat approximately 20 inches from the computer monitor and wore headphones to hear the task instructions and to block out distractor noise. Upon detecting a change, participants used a computer keyboard and internal microphone to make their response.

After signing a standard experimental consent form, participants were given brief instructions of the task before being asked to put on the headphones to hear the instructions in more detail using a standardized video tutorial. After hearing the instructions for a second time, participants were asked whether they had any questions before beginning the experiment.

Each trial began with a fixation cross located in the center of the screen as a way to standardize eye gaze. The fixation cross stayed on the screen for 500 ms, before the presentation of the first image (*A*). This image contained both the background photograph and the target object superimposed. The first image stayed on the screen for 250 ms, which was then followed by a blank white screen ISI of 250 ms. Following the ISI, the second image (*A'*) was presented. This image contained the background-only photograph and remained on the screen for 250 ms. The second image was followed by another blank white screen ISI for 250 ms. The sequence then began again from the first image (*A*) onwards for a potential of 19 cycles (19 seconds) or until the target object had been detected and the keyboard button had been depressed. Upon detection, participants were required to press the spacebar, which terminated presentation of the sequence of visual images. They were then prompted to make a one word response to identify the changing object. The procedure for the flicker task is illustrated in Figure 1.

Each participant was exposed to both context conditions (Context: Congruent vs. Incongruent). For each condition, 70 pairs of images were presented for a total of 140

image pairs. Images from both context conditions were presented to participants in a random order.

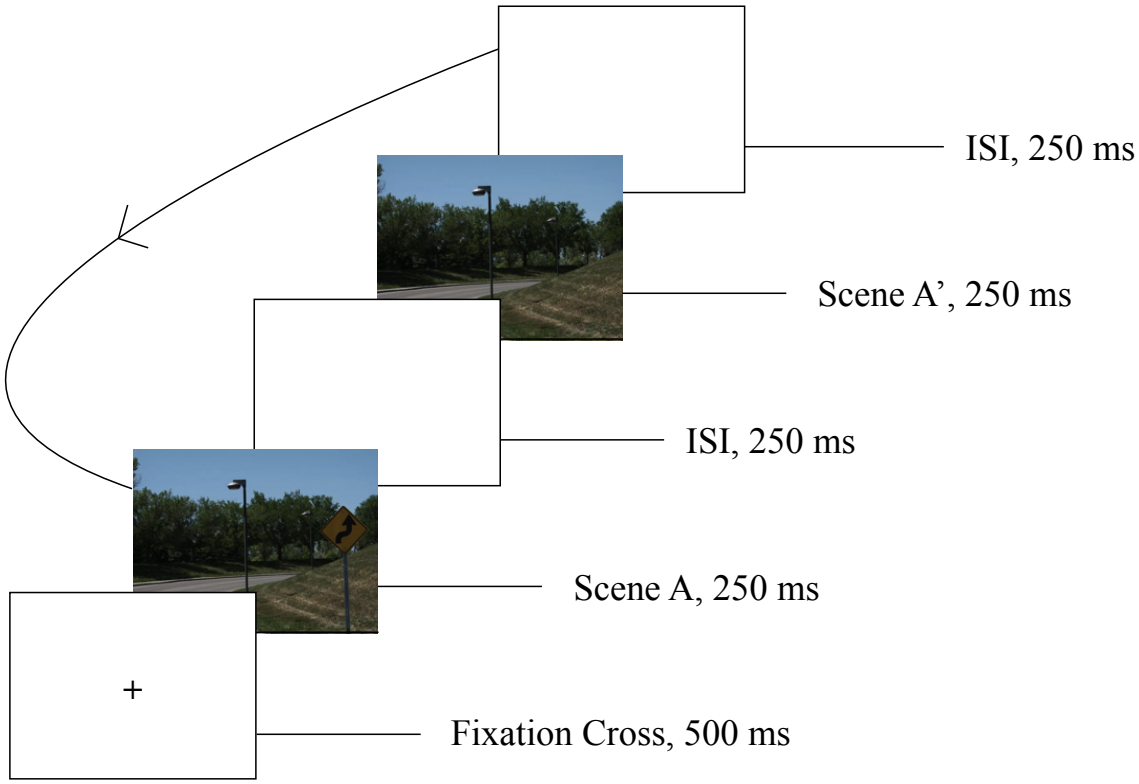

Figure 1. Flicker task used in the first experiment. Each trial began with a fixation cross for 500 ms. The fixation cross was then replaced by the first image (*A*), which contained both the background and target. This image stayed on the screen for 250 ms. The first image was followed by a white inter-stimulus interval (ISI) for 250 ms. Next, the second image (*A'*), containing the background only, was presented for 250 ms. Finally, a second ISI was present for 250 ms. This sequence, from the first image to the second ISI, continued for up to 19 cycles or until a response was made.

Results

There were three key dependent variables analyzed in each of our experiments. Mean response times (RT) were computed using correctly responded to trials only. Misses

were defined as trials in which the change was not detected. Errors were defined as trials in which a change was detected, but an improper object label was given. Mean RTs, miss rates, and error rates were computed separately for the two context conditions for each participant, and these data were then submitted to the analyses described below.

Response Time. Mean RTs for the two context conditions were submitted to a paired sample t-test. Responses were significantly faster for the incongruent context (M = 4833 ms) than for the congruent context (M = 5745 ms), $t(19) = 7.27$, $p < 0.001$. Figure 2 displays the mean RTs for both context conditions collapsed across participants.

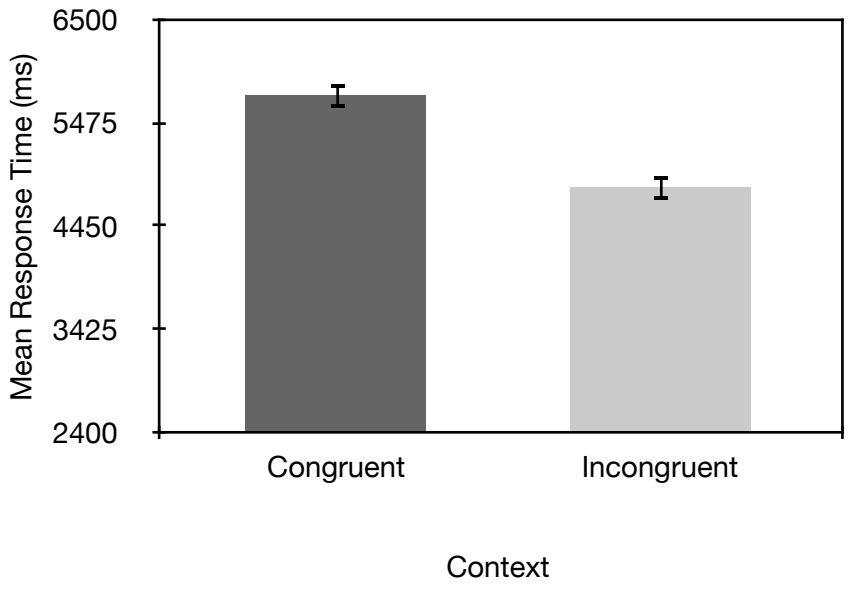

Figure 2. Mean response time (ms) for correctly detected changes in each context category (context: congruent vs. incongruent) for Experiment 1. Response time was measured as the latency between onset of the first presentation of image A (background + target) and initiation of the key press indicating detection of the change. Error bars are standard errors corrected to remove between-subjects variation (Cousineau, 2005).

Misses. Miss rates for the two context conditions were submitted to a paired sample t-test. Misses occurred less often when the change was presented in an incongruent context (0.06) than when it was presented in a congruent context (0.13), $t(19) = 6.39$, $p < 0.001$. That is, the degree of change blindness was larger when the target and background matched contextually. Figure 3 displays the mean proportions of misses for both context conditions collapsed across participants.

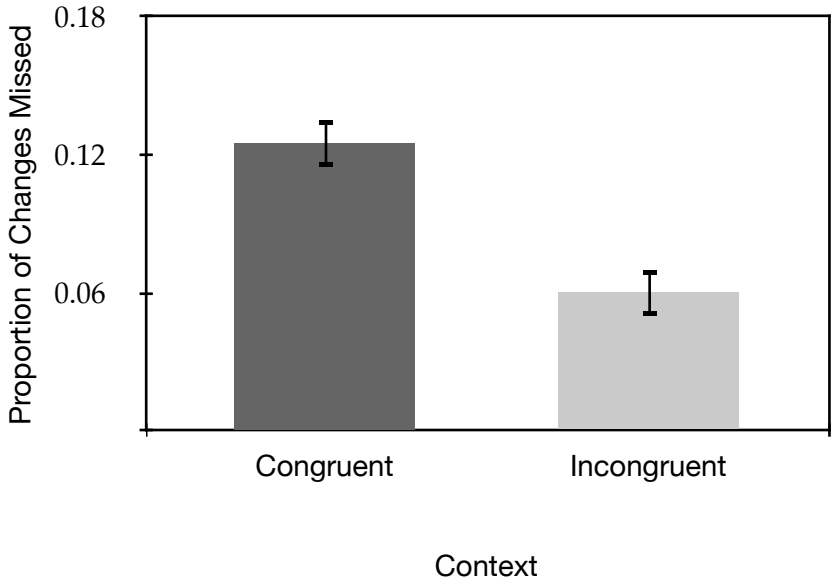

Figure 3. Mean proportion of changes missed in each context category (Context: Congruent vs. Incongruent) for Experiment 1. Misses were defined as trials in which no response was made over the course of the 19 cycles of the flicker task.

Errors. Error rates for the two context conditions were submitted to a paired sample t-test. The error rates for the congruent (.02) and incongruent (.03) context conditions were not significantly different, $t(19) = 1.60$, $p = 0.13$. Figure 4 displays the mean error rates for the two context conditions collapsed across participants.

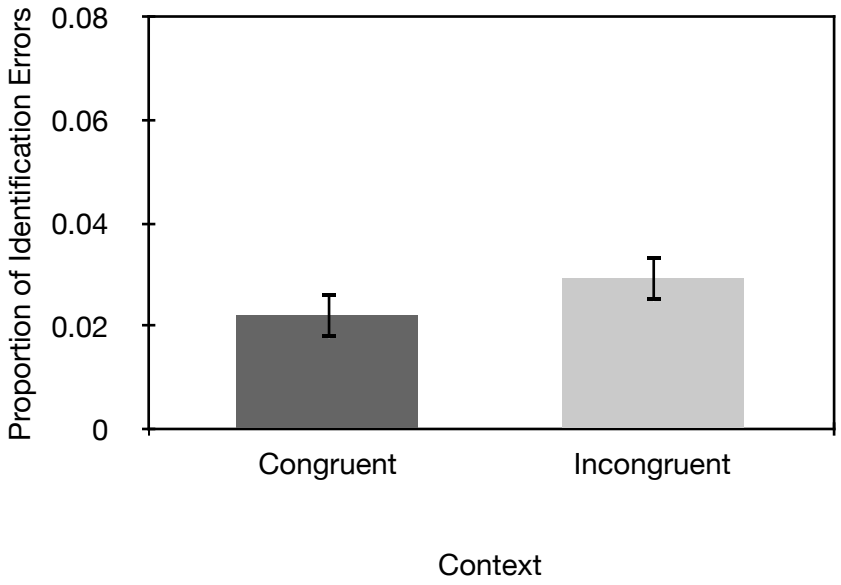

Figure 4. Mean proportion of identification errors in each context category (Context: Congruent vs. Incongruent) for Experiment 1. Identification errors were defined as trials in which a change detection response was made, but an incorrect identification response was given.

Discussion

The goal of this experiment was to establish a change detection procedure that can be used as a standard to measure the context incongruency benefit measured by others (Brockmole & Henderson, 2008; Gordon, 2004; Hollingworth & Henderson, 2007; 2000; Hollingworth et al., 2001; LaPointe, 2011; Pezdek et al., 1989). Indeed, participants responded faster to, and missed fewer, contextually incongruent object changes than contextually congruent object changes. As noted in the Introduction, this result is consistent with the idea that the change detection task introduced a form of attention capture, perhaps guided by rapid gist-based scene processing, that benefits performance for incongruent targets.

Experiment 2

The goal of the present experiment was to examine whether the requirement to detect change necessarily produces a performance benefit for context incongruent targets. If so, then changes to the structure of the change detection procedure should not alter the basic pattern of data as long as the task continues to require detection of a changing object. An alternative view is that the conventional change detection task is ideal for measuring context incongruency benefits because it introduces a long period during which participants repeatedly extract the gist of a visual scene. If attention capture hinges on detecting points of semantic interest that can be derived from gist-based representations, then the extended period during which these gist-based representations are extracted in the typical change detection task may be critical to observe a context incongruency benefit. According to this alternative view, eliminating the temporally extended contribution of gist-based processing to performance might well eliminate the context incongruency benefit, despite the task continuing to require change detection.

As such, the primary aim of this experiment was to assess whether eliminating a temporally extended period during which gist-based processing guides attention capture would also eliminate the context incongruency benefit. Importantly, Rensink et al. (1997) has shown that a temporal interval between successive scenes is critical for producing change blindness in the flicker task. Without a temporal interval between consecutive images, changes can be detected by an easily detectable visual transient at the location of the changing object. As such, removing the temporal interval in our experiment ought to

lead change detection in our task to be driven by rapid detection of a visual transient rather than by a gist-based attention capture process. If this logic holds, then removing the temporal interval should also eliminate the key process that leads to a performance benefit for incongruent targets relative to congruent targets. In its place, we predicted that a context congruency benefit would be observed, supported by processes that facilitate identification of targets that are congruent with the scene in which they are embedded.

To test these predictions, half of the participants in this experiment completed an identical change detection task to that used in Experiment 1. The other half of participants completed a similar task, but with the ISI reduced from 250 ms to 0 ms. We expected that participants in the 250 ms ISI condition would show the same pattern of results as those reported in Experiment 1. That is, we expected these viewers to detect changes to contextually incongruent objects faster than changes to congruent objects and to miss detecting more contextually congruent objects than contextually incongruent objects. In contrast, for participants in the 0 ms ISI condition, we expected changes to congruent objects to be detected faster than changes to incongruent objects. As the 0 ms ISI ought to lead to rapid detection of the changing objects, we expected that there would be too few misses to make meaningful predictions concerning the benefits of context congruency on miss rates.

Method

Participants. Forty undergraduate psychology students from McMaster University, none of whom had participated in Experiment 1, volunteered to participate in exchange for partial course credit.

Stimuli. The stimuli used in the present experiment were the same as those used in Experiment 1.

Materials and Procedure. The same materials as those used in the Experiment 1 were used for the current experiment.

The procedure for the current experiment was also the same as in Experiment 1 with the exception of the introduction of a between-subject manipulation of the ISI. Participants were randomly assigned to either the 250 ms ISI condition or the 0 ms ISI condition. Those who participated in the 250 ms ISI condition were exposed to the same procedure as used in Experiment 1. The procedure for the 0 ms ISI condition was similar, however the blank white ISI was eliminated. Figure 5 illustrates the stimulus presentation parameters of the flicker task for the 0 ms ISI condition.

Within each of the between-subject conditions (ISI: 250 ms vs. 0 ms), participants were exposed to both levels of the within-subject manipulation of context (Context: Congruent vs. Incongruent). For both within-subject conditions, 70 pairs of images were presented for a total of 140 image pairs. Images from the two contextual congruency conditions were presented to participants intermixed in random order.

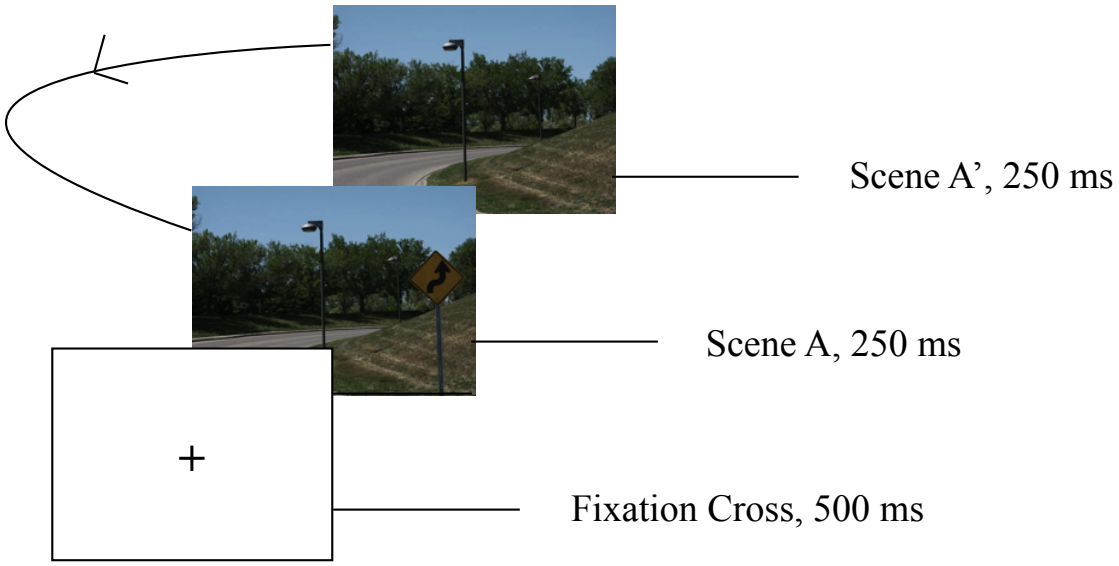

Figure 5. Flicker task used for the 0 ISI condition in Experiments 2 and 3. Each trial began with a fixation cross, which remained on the screen for 500 ms. Next, the first image (*A*) was presented for 250 ms. This image included both the background and the target object. The first image was followed immediately by the second image (*A'*) for 250 ms. The second image contained the background only. This sequence, from the first image to the second, continued for a total of 19 cycles or until a response was made using the keyboard.

Results

The key dependent variables in this experiment were the same as in Experiment 1. Mean RTs, miss rates, and error rates were computed for each condition defined by the factorial combination of the context (congruent vs. incongruent) and ISI (0 ms vs 250 ms) variables, separately for each participant. These data were submitted to mixed factor analyses of variance that treated context as a within-subject factor and ISI as a between-subjects factor.

Response Time. There was a significant main effect of ISI, F (1, 38) = 117.38, MSE = 829627.87, $p < 0.001$, with changes to targets in the 0 ms ISI condition (M = 2956 ms) being detected significantly faster than changes to targets in the 250 ms ISI condition (M = 5163 ms). More important, we also found a significant interaction between context and ISI, F (1, 38) = 37.96, MSE = 71520.30, $p < 0.001$.

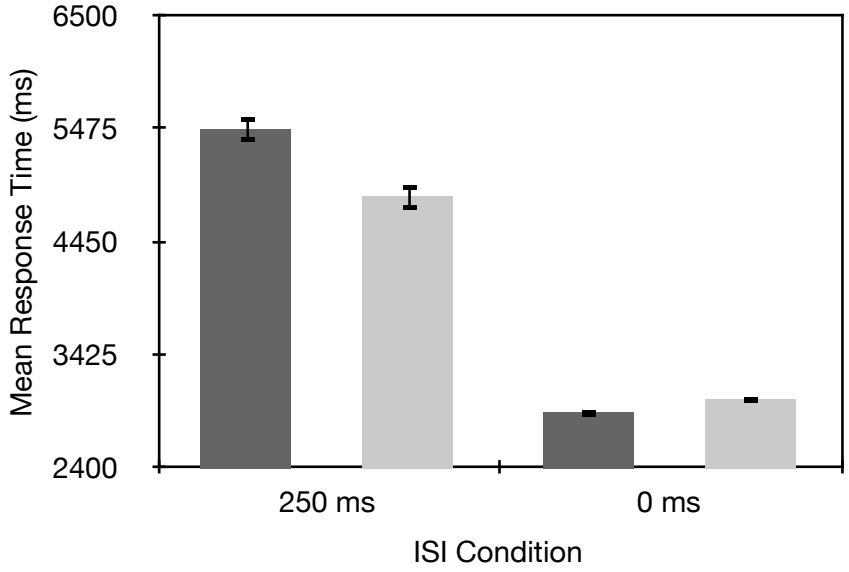

Figure 6. Mean response time (ms) for correctly detected changes in each context category (context: congruent vs. incongruent) for both ISI conditions (ISI: 250 ms vs. 0 ms) in Experiment 2 where participants were required to identify the changing object.

This interaction was examined further by analyzing the effect of context separately for the 250 ms and 0 ms ISI conditions. For the 250 ms ISI condition, participants were faster at detecting the target for incongruent trials (M = 4855 ms) than for congruent trials (M = 5470 ms), $t(19) = 5.34$, $p < 0.001$. In contrast, for the 0 ms ISI condition, participants were faster at detecting the target for congruent trials (2895 ms) than for incongruent trials

(M = 3017 ms), $t(19) = 3.76$, $p = 0.001$. The mean RTs for each condition, collapsed across participants, is presented in Figure 6.

Misses. There was a significant main effect of ISI, F (1, 38) = 140.51, MSE = 0.002, $p < 0.001$, with a larger proportion of misses for the 250 ms ISI condition (M = 0.11) than for the 0 ms ISI condition (M = 0.003). Once again, there was also a significant interaction between context and ISI, F (1, 38) = 79.02, MSE = 0.00, $p < 0.001$. To examine this interaction further, the effect of context was examined separately for the 250 ms and 0 ms ISI conditions.

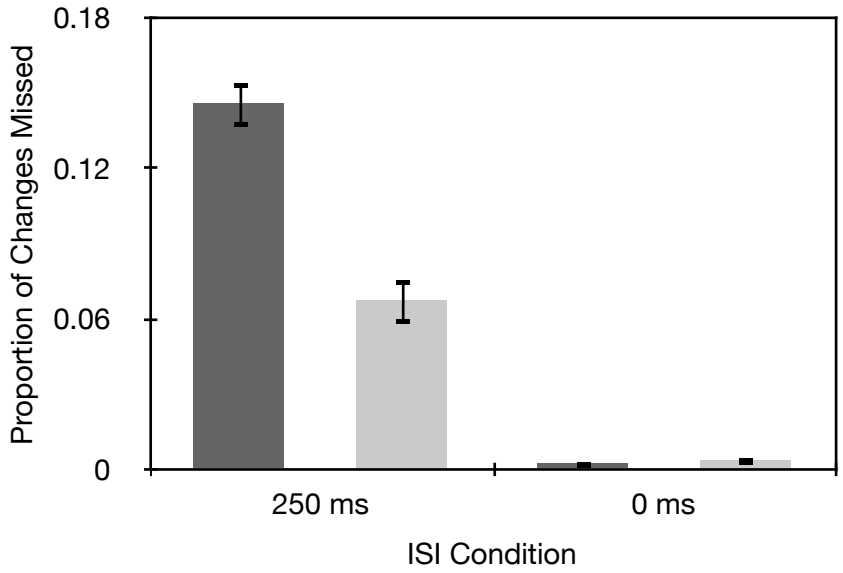

Figure 7. Mean proportion of changes missed in each context category (context: congruent vs. incongruent) for both ISI conditions (ISI: 250 ms vs. 0 ms) in Experiment 2 where participants were required to identify the changing object.

For the 250 ms ISI condition, participants missed fewer changes to the scenes for incongruent trials (M = 0.07) than for congruent trials (M = 0.15), $t(19) = 8.78$, $p < 0.001$.

In contrast, for the 0 ms ISI condition, there was no difference in miss rates between the congruent (M = 0.002) and incongruent (M = 0.004) conditions, $t(19) = 1.45$, $p = 0.16$. The mean miss rates for each condition, collapsed across participants, are displayed in Figure 7.

Errors. The only significant effect in this analysis was the interaction between context and ISI, $F(1, 38) = 4.49$, $MSE = 0.00$, $p = 0.04$. To examine this interaction further, the effect of context was analyzed separately for the 250 ms and 0 ms ISI conditions. For the 250 ms ISI, error rates in the congruent (M = 0.03) and incongruent (M = 0.04) conditions did not differ significantly, $t(19) = 1.41$, $p = 0.18$. However, for the 0 ms ISI, more errors were made in the incongruent (M = 0.04) than in the congruent (M = 0.01) condition, $t(19) = 4.81$, $p < 0.001$. Mean error rates for each condition, collapsed across participants, are displayed in Figure 8.

Discussion

The primary goal of the second experiment was to determine whether detecting change in itself is enough to produce a context incongruency benefit due to an attention capture process directed towards points of semantic interest in a scene. According to this explanation we should continue to see a context incongruency benefit when detecting change despite eliminating a protracted period of gist-based change detection. The alternative view is that it is the extended temporal window of gist-based change detection that is key to finding a context incongruency benefit. According to this view, removing the protracted period of gist-based change detection ought to remove the context incongruency benefit, and perhaps uncover in its place a context congruency benefit.

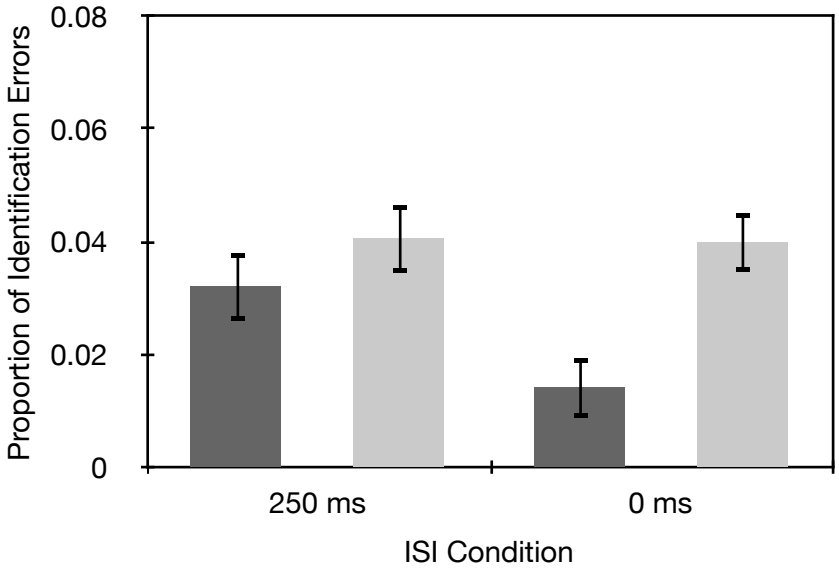

Figure 8. Mean proportion of identification errors in each context category (context: congruent vs. incongruent for both ISI conditions (ISI: 250 ms vs. 0 ms) in Experiment 2 where the participants were required to identify the changing object.

To test this issue, we reduced the ISI between image presentation to 0 ms. Indeed, the context incongruency benefit observed in the 250 ms ISI condition reversed to a context congruency benefit in the 0 ms ISI condition. The errors participants made in identifying targets objects in the 0 ms ISI condition were consistent with the idea that change detection performance in this condition was predominantly controlled by target identification processes that favor processing of context congruent targets. In particular, participants made more errors in identification when the object was presented as part of an incongruent scene.

Experiment 3

In Experiment 2, we introduced a manipulation designed to reduce the contribution of gist-based attention capture to change detection by reducing the ISI to 0 ms. Under these conditions, we uncovered a context congruency benefit, in terms of both response time and object identification accuracy. The goal of the present experiment was to minimize the contribution of object identification processes to performance on the change detection task. The rationale here was to minimize the contribution of a process presumed to benefit from context congruency (i.e., identification). In the absence of such a process, there ought to be large change detection effects favoring context incongruency for the 250 ms ISI condition. More important, minimizing the contribution of target identification processes to change detection performance ought to eliminate the context congruency benefit observed for the 0 ms ISI condition in Experiment 2. To this end, rather than asking participants to identify the changing object to confirm that they had correctly detected which object was changing, participants were instead instructed to locate the area within each image in which the target object appeared and disappeared.

Method

Participants. Forty-two undergraduate psychology students from McMaster University, none of whom participated in Experiments 1 or 2, volunteered to participate in exchange for partial course credit.

Stimuli. The same stimuli used in Experiments 1 and 2 were used for the current experiment.

Materials and Procedure. The same materials as those used in the Experiments 1 and 2 were used for the current experiment. The procedure for this experiment was very similar to that used in Experiment 2. As in Experiment 2, upon detecting the changing object, participants were asked to press the spacebar, which terminated the presentation of the sequence of alternating images. However, a 9-box (3 x 3) grid was then presented in place of the images. Participants were then prompted to identify verbally in which of the boxes the change had taken place. In other words, the participants' task was to localize the changing object rather than to identify the changing object. As in Experiment 2, context (congruent/incongruent) was manipulated within-subject, while ISI (0 ms/250 ms) was manipulated between-subjects.

Results

The key dependent variables in this experiment were the same as in Experiments 1 and 2. As in Experiment 2, mean RTs, miss rates, and error rates were computed for each condition defined by the factorial combination of the context (congruent vs. incongruent) and ISI (0 ms vs. 250 ms) variables, separately for each participant. These data were submitted to mixed factor analyses of variance that treated context as a within-subject factor and ISI as a between-subjects factor.

Response Time. There was a significant main effect of ISI, F (1, 40) = 1518.56, MSE = 886986.96, $p < 0.001$, with changes to targets in the 0 ms ISI condition (M = 2883 ms) being detected significantly faster than changes to targets in the 250 ms ISI condition

(M = 5126 ms). More important, we also found a significant interaction between context and ISI, F (1, 40) = 67.85, MSE = 62212.77, $p < 0.001$.

This interaction was examined further by analyzing the effect of context separately for the 250 ms and 0 ms ISI conditions. For the 250 ms ISI condition, participants were faster at detecting the target for incongruent trials (M = 4706 ms) than for congruent trials (M = 5545 ms), $t(20) = 7.91$, $p < 0.001$. For the 0 ms ISI condition, participants were also faster at detecting the target for incongruent trials (2853 ms) than for congruent trials (M = 2911 ms), $t(20) = 2.34$, $p = 0.03$, although this difference was smaller than for the 250 ms ISI. The mean RTs for each condition, collapsed across participants, is presented in Figure 9.

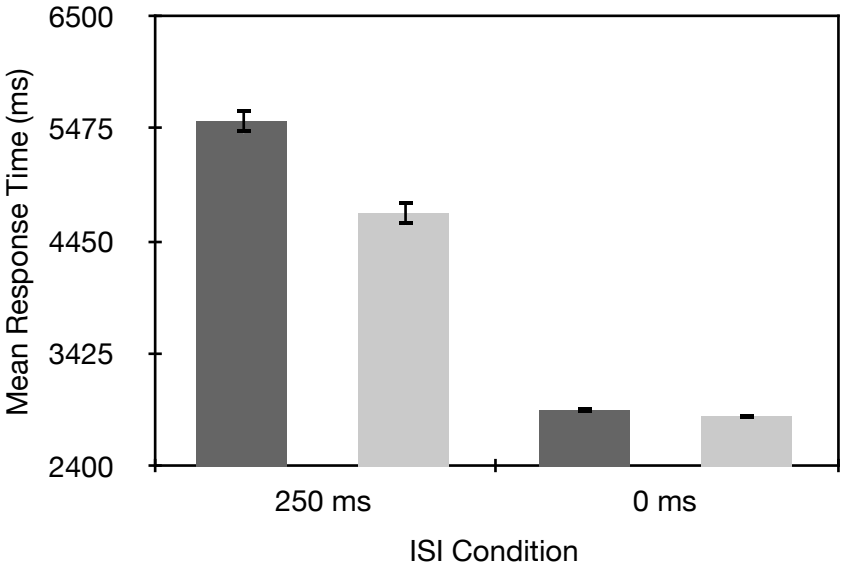

Figure 9. Mean response time (ms) for correctly detected changes in each context category (context: congruent vs. incongruent) for both ISI conditions (ISI: 250 ms vs. 0 ms) in Experiment 3 where participants were required to localize the changing object.

Importantly, we also compared the mean response times for each context condition for the 0 ms ISI condition across Experiments 2 and 3. We found a significant interaction between experiment and context, F (1, 39) = 19.31, MSE = 139409.71, $p < 0.001$.

Misses. There was a significant main effect of ISI, F (1, 40) = 113.95, MSE = 0.002, $p < 0.001$, with a larger proportion of misses for the 250 ms ISI condition ($M = 0.11$) than for the 0 ms ISI condition ($M = 0.002$). Once again, there was also a significant interaction between context and ISI, F (1, 40) = 86.04, MSE = 0.00, $p < 0.001$. To examine this interaction further, the effect of context was examined separately for the 250 ms and 0 ms ISI conditions.

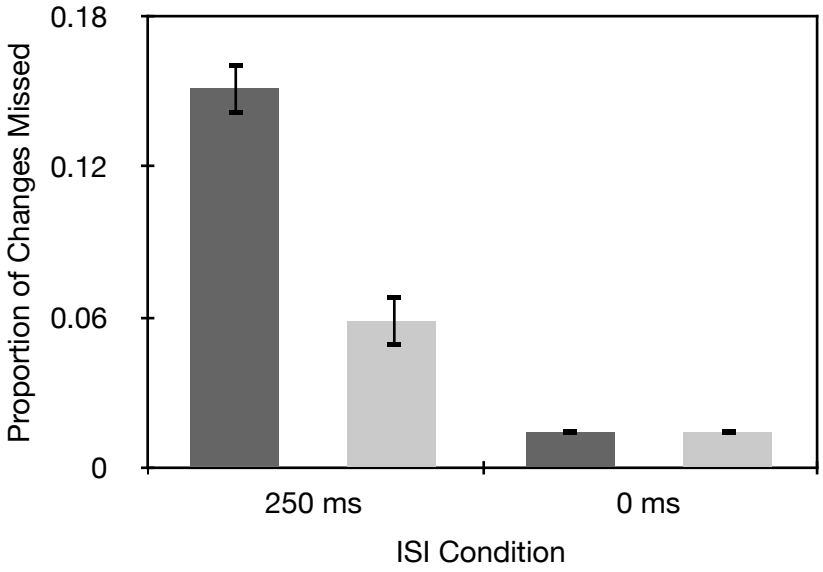

Figure 10. Mean proportion of changes missed in each context category (context: congruent vs. incongruent) for both ISI conditions (ISI: 250 ms vs. 0 ms) in Experiment 3 where participants were required to localize the changing object.

For the 250 ms ISI, participants missed fewer changes to the scenes for incongruent trials ($M = 0.06$) than for congruent trials ($M = 0.15$), $t(20) = 9.55$, $p < 0.001$. In contrast, for the 0 ms ISI, participants missed marginally more changes to the scenes for incongruent trials ($M = 0.003$) than for congruent trials ($M = 0.001$), $t(20) = 1.83$, $p = 0.08$. The mean miss rates for each condition, collapsed across participants, are displayed in Figure 10.

Errors. There were no significant effects in the analysis of the error rates. Mean error rates for each condition, collapsed across participants, are displayed in Figure 11.

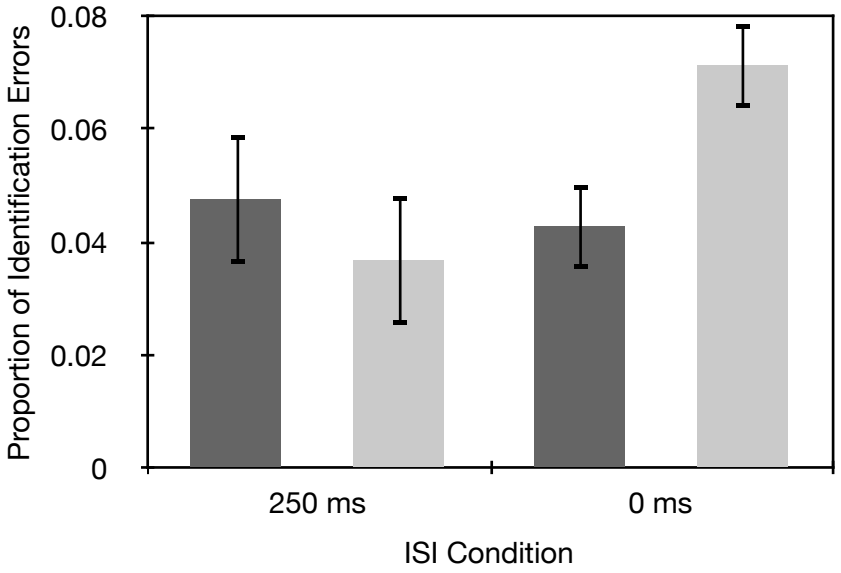

Figure 11. Mean proportion of localization errors in each context category (context: congruent vs. incongruent) for both ISI conditions (ISI: 250 ms vs. 0 ms) in Experiment 3 where participants were required to localize the changing object.

Discussion

The goal of this experiment was to minimize the contribution of object identification processes to change detection performance by requiring participants to localize rather than identify the changing object. The key prediction was that this manipulation ought to eliminate the context congruency benefit that had been observed for the 0 ms ISI condition in Experiment 2. Indeed, there was no evidence of a context congruency benefit in the 0 ms ISI condition of the current experiment. Rather, there was a significant context incongruency benefit in this condition. Although this result was not anticipated, it perhaps suggests that the attention capture mechanism responsible for context incongruency benefits can express itself in performance on the same time scale as the influence of visual transients, on at least some small proportion of trials. In any case, the opposite effects of context congruency in the 0 ms ISI condition across Experiments 2 and 3 is consistent with the focal theme of this article. In particular, attention capture processes that help participants detect and localize the changing target on the one hand, and target identification processes that allow participants to identify the changing target on the other hand, produce opposite effects of context congruency in change detection tasks.

General Discussion

The broad purpose of the present study was to examine the processes that contribute to context congruency effects in change detection performance. The results of Experiment 1 replicated the common finding that change detection is often more efficient for context incongruent than context congruent targets (Brockmole & Henderson, 2008; Hollingworth

& Henderson, 2000; LaPointe, 2011; Loftus & Mackworth, 1978). This finding is interesting in light of the fact that congruent contexts contain a host of cues as to what types of objects are likely to be encountered, their probable location, and their interaction with other objects. In fact, there are a number of studies in the visual search literature that describe an advantage in search for targets in congruent relative to incongruent contexts (Davenport & Potter, 2004; Henderson et al., 1999; Kelley et al., 2003; Neider & Zelinsky, 2006; Palmer, 1975). As such, more efficient performance for incongruent than congruent targets in change detection tasks must reflect the operation of a process that differs from those that guide performance in a range of other search tasks.

A central question addressed by Experiments 2 and 3 was whether the context incongruency benefit often observed in change detection tasks is necessarily linked to the task goal of detecting change. If this were the case, then any and all tasks requiring change detection ought to produce the same effect. An alternative view is that whether one finds a context incongruency benefit is likely to depend on the processes that guide shifts of attention to the location of the change, and that the prolonged period of search for a change in many change detection tasks introduces an opportunity for gist-based attention capture processes to guide shifts of attention to context incongruent targets. By this view, if shifts of attention were instead guided by detection of a visual transient, then a context incongruency benefit should not occur.

The results of Experiments 2 and 3 were consistent with this latter idea, and in particular with the idea that context congruency effects in change detection can be affected

by at least two distinct processes; gist-based attention capture processes that favour shifts of attention to context incongruent targets, and object identification processes that favour the identification of context congruent targets. In Experiment 2, we attempted to focus on the second of these two processes. If the interaction of semantic information contained in the target and background favours shifts of attention to context incongruent targets, then the contribution of this process ought to be minimized by reducing the ISI between successive scene presentations from 250 ms to 0 ms. With a 0 ms ISI, shifts of attention to the changing target ought to be guided by detection of visual transients rather than by gist-based attention capture. Indeed, when the ISI was reduced to 0 ms, change detection performance was faster for congruent targets than for incongruent targets, and there were fewer identification errors for congruent targets than for incongruent targets. Together, these results are consistent with the view that change detection performance is affected by two processes that are influenced in opposite ways by context congruency.

The purpose of Experiment 3 was again to change the processing requirements of the change detection task, in this case with the aim of reducing the contribution of object identification processes that benefit from congruent contexts. To do this, we asked participants to detect the change and then to localize, rather than identify, the changing object. In line with our prediction, regardless of whether the ISI was 0 ms or 250 ms, change detection performance was now faster for incongruent trials than for congruent trials. The fact that change detection performance changed qualitatively for the 0 ms ISI condition across Experiments 2 and 3 was particularly noteworthy. When the task required

identification (Experiment 2) change detection was faster for congruent targets. In contrast, when the task required localization (Experiment 3), change detection was faster for incongruent targets.

This set of results is consistent with several other recent studies on the influence of context in scene and object perception. For example, Sun et al. (2011) presented scenes to participants for 20 ms in a go/no-go categorization task, where participants were to indicate whether the scene contained an animal. Participants were significantly faster and more accurate at detecting an animal if it was accompanied by a congruent context. Moreover, the onset latency of a frontal ERP marker of object recognition was earlier for target object appearing with a congruent scene context than with a phase randomized scene context. These results fit well with the present observation that context congruency facilitates object identification processes (see also Vo & Henderson, 2011; Vo & Schneider, 2010).

Perhaps most similar to the current study are results from LaPointe (2011). In an experiment very similar to the present Experiment 1, the author observed more efficient change detection performance for context incongruent targets. However, in a subsequent experiment a gaussian filter was placed over the target object, with the goal of disrupting its semantic informativeness. If the performance benefit for contextually incongruent objects is driven by the interaction of the semantic information of the target and the background (i.e., semantic points of interest), then under these circumstances the advantage for context incongruent targets should not be found. Indeed, when the target objects were

blurred, there was no difference in change detection performance for context congruent and context incongruent targets.

Although we have adopted the idea that the advantage for context incongruent targets in many change detection tasks stems from shifts of attention to points of semantic informativeness, it is worth noting that there are alternatives to the view that this attention capture process hinges on semantic processing. In particular, Gordon (2006) noted a distinction between a semantic conflict interpretation, like that forwarded here, and an alternative in which attention capture is driven by the difficulty of perceptual identification of objects in the scene. If context congruent targets are identified more easily than are context incongruent targets, then attention shifts to objects that are difficult to identify would produce the type of context incongruency benefits of interest in the present study. Gordon (2006) offered support for this view in a study that examined lexical decision performance for letter strings following presentation of visual scenes. A key result in this study was that lexical decision targets that matched an item in the preceding scene that was incongruent with the scene context produced a null priming effect. At the same time, a negative priming effect was observed for lexical decision items that matched objects in the preceding scene that were consistent with the scene context. Gordon (2006) suggests that attention shifts to local areas of the scene in which perceptual processing is difficult (i.e., context incongruent objects), together with suppression of context congruent objects in the scene, would explain both of these priming results without recourse to the idea that attention shifts are driven by semantic conflict. Our sense is that the present results are

equally consistent with the semantic conflict and perceptual processing difficulty interpretations of attention capture by context incongruent targets. As the present experiments aimed at teasing apart separate attention capture and identification components of change detection performance, the processing basis of the attention capture component must await further research.

In summary, we have argued here that varying task demands of a change detection task selectively modulates the extent to which performance depends on two processes that are affected in opposite ways by context congruency. Whereas a gist-based attention capture process may pull attention toward targets objects that are incongruent with the surrounding scene context, object identification processes are more efficient for target objects that are congruent with the surrounding scene context. A key conclusion that we draw is that change detection tasks do not invariably result in more efficient performance for context incongruent than context congruent targets. Rather, the effect of context congruency on change detection performance hinges on the relative contributions of at least two processes that are free to vary quite widely across many variants of change detection tasks.

Chapter 3: Semantically Incongruent Objects Attract Eye-Gaze when Viewing Scenes for Change

LaPointe, M. R. P., & Milliken, B. (2016).

Visual Cognition, Online first publication.

DOI: 10.1080/13506285.2016.1185070

Preface

The experiments in the previous chapter demonstrate that multiple processes underlie complex scene perception and the relative weighting of these processes can be manipulated in such a way as to produce congruency benefits or incongruency benefits within a task. Yet, a pressing question in the literature revolves around the nature of attention allocation when these processes are weighted in favour of congruency costs. Specifically, two competing hypotheses have been proposed to account for congruency costs (Hollingworth & Henderson, 2000). The attention attraction hypothesis assumes the conflict in meaning between an incongruent object and the scene context attracts attention. The attention disengagement hypothesis assumes attention is dispersed about the scene randomly, but lingers on incongruent objects for additional processing. The experiment reported in Chapter 3 was designed to identify the nature of attention allocation when

processing is weighted in favour of incongruency benefits when perceiving complex scenes. To do so, a typical change detection task (Rensink, O'Regan, & Clark, 1997) was used while also monitoring participants' eye movements. The results demonstrate a replication of incongruency benefits when performing a change detection task. Moreover, across a number of eye movement measures, we report evidence that attention is attracted to semantically incongruent objects, but fail to find any evidence that attention lingers on these objects.

Abstract

Past research has shown that change detection performance is often more efficient for target objects that are semantically incongruent with a surrounding scene context than for target objects that are semantically congruent with the scene context. One account of these findings is that attention is attracted to objects for which the identity of the object conflicts with the meaning of the scene, perhaps as a violation of expectancies created by earlier recruitment of scene gist information. An alternative account of the performance benefit for incongruent objects is that attention is more apt to linger on incongruent objects, as perhaps identifying these objects is more difficult due to conflicting information from the scene context. In the current experiment, we present natural scenes in a change detection task while monitoring eye-movements. We find that eye-gaze is attracted to these objects relatively early during scene processing.

Introduction

Efficient cognitive functioning relies heavily on a structured and stable environment (Gibson, 1979). For example, the structure inherent in the context of a visual scene allows attention to be directed to areas of the scene that are informative (Antes, 1974; Buswell, 1935; Hollingworth & Henderson, 2000; Mackworth & Morandi, 1967; Stirk & Underwood, 2007; Yarbus, 1967), interesting (Kelley, Chun, & Chua, 2003; Rensink, O'Regan, & Clark, 1997; Turatto, & Galfana, 2000), or goal-relevant (Hayhoe & Ballard, 2005; Henderson, Brockmole, Castelhano, & Mack, 2007). Consider a scenario of driving along a busy highway – knowledge of the context, including what types of objects to expect as well as their probable locations, allows efficient allocation of attention to areas of the scene that promote the safe operation of the vehicle. At the same time, prior knowledge of visual contexts allows goal-irrelevant, yet "typical", areas of a scene (e.g., lamp posts) to be spared focal attention.

Prior knowledge of visual contexts, however, can also promote attention allocation towards "atypical" areas of a scene (Bonitz & Gordon, 2008; Brockmole & Henderson, 2008; Gordon, 2004; 2006; Hollingworth & Henderson, 2000; 2007; Hollingworth, Williams, & Henderson, 2001; LaPointe, Lupiáñez, & Milliken, 2013; Loftus & Mackworth, 1978; Pezdek, Whetstone, Reynolds, Askari, & Dougherty, 1989). Consider again driving along a busy highway – encountering a deer amongst the traffic is likely to gain the focus of attention. Hollingsworth and Henderson (2000) describe objects that are semantically incongruent with a scene context as being particularly informative because

these objects contain little overlapping information with other objects in the scene. In contrast, semantically congruent objects are less informative because they contain information that is redundant with other objects in the scene. Of particular interest for the present study is when and how semantically incongruent objects attract attention.

Semantic Congruency Effects in Scene Processing: Mixed Findings

The attraction of attention to semantically incongruent objects is intriguing due to the mixed set of results that both support and fail to support the existence of this phenomenon. Some empirical work has shown that semantically incongruent objects attract attention early during scene processing. Take for example an experiment by Loftus and Mackworth (1978), who presented participants with complex line drawings containing target objects that were either semantically congruent or incongruent with the scene context. The scenes were presented for four seconds and participants were asked to use that time to investigate the scene in preparation for a later scene recognition memory test. In this case, participants fixated incongruent targets significantly faster than congruent targets. In fact, participants were more likely to fixate the incongruent targets within the first two fixations of scene viewing, and the saccades towards these objects were large compared to those shifting towards congruent targets. These results have garnered attention in the scene processing literature because they suggest not only that semantically incongruent objects are fixated earlier than semantically congruent objects, but also that these objects are processed semantically, at least to some degree, in the visual periphery prior to foveal processing.

Replicating the eye-movement patterns reported by Loftus and Mackworth (1978), however, has proven to be difficult (D'Graef, Christiaens, & d'Ydewalle, 1990; Friedman, 1979; Gareze & Findlay, 2007; Henderson, Weeks, & Hollingworth, 1999; Rayner, Castelhano, & Yang, 2009; Vo & Henderson, 2009; 2011). For example, using a similar task, Henderson et al. (1999) presented complex line drawings for 15 seconds each, also under threat of a scene recognition memory test. These researchers, however, reported no difference in the probability of immediate fixation on the target objects, nor did they find any difference in the number of fixations prior to fixating the target object. Moreover, in direct contrast to the result reported by Loftus and Mackworth, the size of the initial saccade toward the target object was larger for congruent than for incongruent objects. These results do not indicate an early, peripheral processing advantage for incongruent targets, as suggested by Loftus and Mackworth. Henderson et al., however, did show that participants fixated the incongruent objects longer and more often, suggesting additional processing of these objects is required during complex scene viewing. Taken together, and despite using similar tasks, the results reported in these two studies provide entirely different answers to the question of whether semantically incongruent objects attract attention during early scene processing.

In more recent years, researchers have probed the early processing of semantically incongruent objects by using a different set of tasks. Take for example an experiment conducted by Vo and Henderson (2011), wherein the authors attempted to isolate the contribution of contextual information to only the very early moments of scene viewing.

The key idea was that if semantically incongruent objects attract attention early during scene processing, then preferential processing for these objects should manifest after only a short preview of the scene. Here, each trial began with a 250 ms preview of a computer generated complex scene, followed by a 50 ms mask. Participants were then given a word, which remained on the screen for 1500 ms, that described a target object that was either semantically congruent or incongruent with the scene. A refixation cross was then presented for 500 ms, followed by re-presentation of the search scene. The task was simply to move the eyes to the location of the target object in the final search scene. What is particularly interesting about this experiment is that participants' view during the final search scene was restricted to a 5° diameter window that moved in accordance with their eye movements. Each search scene remained on the screen for 15 seconds or until the target object was found. Given the scene context was presented for just 250 ms at the beginning of the trial, its influence was restricted to only the very early moments of scene processing. Vo and Henderson found no difference in the time at which the first saccade was made, the size of the first saccade, or the size of the saccade entering the target area across the congruent and incongruent context conditions. In this case, however, participants took less time to respond on congruent trials, less time to fixate the congruent targets, and made fewer fixations before landing on congruent targets compared with incongruent targets. These results are in line with those reported by Henderson et al. (1999) and suggest that eye gaze is not attracted by semantically incongruent objects early during scene processing.

Using another task, Gordon (2004) offers yet different results. Gordon aimed to isolate the influence of scene context to the early moments of scene processing by presenting participants with complex line drawings that contained both a semantically congruent and incongruent target for either 53 or 147 ms. Following the scene, a mask containing a probe appeared for 107 ms. Participants were required to indicate whether the probe was a percent sign or an ampersand. Crucially, the probe appeared in the same location as one of the two target objects. In the 147 ms scene presentation condition participants were more likely to attend and correctly identify the probe when it was located in the same area as the semantically incongruent target. Gordon interpreted these results as evidence that viewers preferentially orient to objects that semantically conflict with the scene meaning and that this orienting of attention occurs within 147 ms of scene processing, but not as early as 53 ms. These results seem at odds with the results of the two studies described above (Henderson et al., 1999; Vo & Henderson, 2011; see also Friedman, 1979). However, they seem to corroborate, at least in part, the results reported early on by Loftus and Mackworth (1978), who suggested that semantically incongruent objects are attended early and preferentially during scene processing.

Semantic Congruency Effects in Change Detection

The results summarized above suggest no clear answer to whether semantically incongruent objects are processed differently than semantically congruent objects when viewing a complex scene. Using a free viewing paradigm, Loftus and Mackworth (1978) report evidence in favour of early attraction of attention to semantically incongruent

objects, while Henderson et al. (1999) do not. Moreover, in attempts to isolate the influence of scene context to a glance, experiments by Vo and Henderson (2011) and Gordon (2004) also produced contradictory results.

On the other hand, there is one task that has produced performance benefits for semantically incongruent objects reliably and robustly; change detection using brief and repeated image presentation (Bonitz & Gordon, 2008; Hollingworth & Henderson; 2000; 2007; Hollingworth, Williams, & Henderson, 2001; LaPointe et al., 2013). Using the flicker paradigm (Rensink, O'Regan, & Clark, 1997), two images are presented for a brief amount of time, one after the other and separated by a blank screen with equally short exposure. The first image contains both the background and the target object, while the second contains the background only. The goal of the participant is to detect which object is changing from one image presentation to the next as quickly and accurately as possible. This task may place a strong emphasis on the early stages of scene processing by using brief and repeated stimulus presentation, possibly isolating the stage at which processing shifts from the broad extraction of contextual information in the scene to the identification of local objects (Biederman, 1981; Castelhano & Henderson, 2008; Fei-Fei, Iyer, Koch, & Perona, 2007; Henderson & Hollingworth, 1999; Intraub, 1981; Oliva & Schyns, 1997; Potter, 1975; 1976; Sampanes, Tseng, & Bridgeman, 2008; Schyns & Oliva, 1994; VanRullen & Thorpe, 2001).

Hollingworth and Henderson (2000) used the flicker paradigm to investigate change detection performance for target objects that were semantically congruent and incongruent

with the context in which they were embedded. In this case, participants viewed images of complex line drawings, similar to those used by Henderson et al. (1999). The alternating images were presented for 250 ms each, interleaved by blank screens presented for 80 ms each. On any given trial, the changing target object could be semantically congruent or incongruent. In this case, participants were significantly faster at detecting changes to incongruent relative to congruent targets. These results are in line with the idea that viewers attend to semantically incongruent objects early during scene processing (Loftus & Mackworth, 1978).

Conflicting Accounts of Change Detection

Hollingworth and Henderson (2000) urged caution in interpreting these change detection results, as early attention capture by semantically incongruent objects is not the only plausible account. Nonetheless, it is possible that as a viewer moves from the rapid recruitment of context information, shown to occur within 100 ms or less (Biederman, 1981; Castelhano & Henderson, 2008; Fei-Fei, Iyer, Koch, & Perona, 2007; Henderson & Hollingworth, 1999; Intraub, 1981; Oliva & Schyns, 1997; Potter, 1975; 1976; Sampanes, Tseng, & Bridgeman, 2008; Schyns & Oliva, 1994), to the recruitment of object information, shown to occur at 150 ms or later (VanRullen & Thorpe, 2001), a conflict between a semantically incongruent object and the context attracts attention for further processing. Hollingworth and Henderson have labeled this account the *attention attraction hypothesis*, wherein early attraction to incongruent objects produces fast change detection. These authors, however, offer an alternative account, which they term the *attention*

disengagement hypothesis. According to this alternative view, when looking for a changing object viewers randomly sample areas of the scene, and when attention lands on a semantically incongruent object it lingers for additional processing, perhaps in an attempt to identify the oddball object (LaPointe et al., 2013; Palmer, 1975). The longer the time spent processing the incongruent object, the more likely the next scene presentation will occur, thereby increasing the likelihood that a change will be detected.

The Current Experiment

The results summarized above offer conflicting results, both within and across tasks, in terms of whether semantically incongruent objects garner early processing relative to semantically congruent objects. These competing results and interpretations lead to two pressing questions: (1) If semantic conflict attracts attention, why do benefits for incongruent objects appear in some tasks and procedures but not others? (2) When benefits for incongruent objects do appear consistently within a task (e.g., change detection), can they be attributed to processes other than attention attraction? In the current experiment, we attempt to answer the second of these two questions by using a change detection paradigm, a task that has reliably shown performance benefits for semantically incongruent objects, and using eye-movement measures to monitor early exploration of natural scenes. In particular, we aimed to evaluate whether eye-gaze, and presumably focal attention, is attracted to incongruent objects early during scene processing in a change detection task. We also aimed to evaluate whether eye-gaze lingers on incongruent target objects during scene processing in a change detection task, as suggested by the attention disengagement

hypothesis. It is important to note that the attention attraction and attention disengagement hypotheses are not mutually exclusive; eye-gaze could be drawn quickly to incongruent objects during early scene processing and linger on these objects for further processing.

Method

Participants. Nineteen undergraduate psychology students (11 female) from McMaster University ranging in age from 17 to 26 years (M = 18.5, SD = 1.93) volunteered to participate in exchange for partial course credit. All participants self-reported normal or corrected-to-normal vision. Four participants were excluded from the final analyses due to difficulty calibrating the eye-gaze tracker prior to the beginning of the experiment, leaving a total of 15 participants (9 female) ranging in age from 17 to 26 years (M = 18.67, SD = 2.16).

Apparatus and Stimuli. The images used in the current experiment were a subset of the images used by LaPointe et al. (2013) and were created from photographs taken in Brisbane, Australia. In total, 126 image pairs were used, with one image of each pair serving as the background-only image (*A'*) and the other serving as the background-plus-target image (*A*). The background-only images were created from photographs of natural scenes that differed on a number of measures, including degree of complexity, number of objects, lighting conditions, landscapes, and scene contexts (e.g., kitchen). The background-plus-target images were created by digitally superimposing a target object from a separate photograph onto a copy of the background-only image. Half of the target objects were placed on background images that were semantically congruent and half were

placed on background images that were semantically incongruent. In all, 63 image pairs were presented in the congruent condition and 63 were presented in the incongruent condition.

In the congruent condition, target objects were placed in background images that depicted a context in which the target object would normally be found. In this condition, target objects were placed in a naturally fitting and physically plausible area of the scene (e.g., a sign post next to a street, bottom left corner of the image). In the incongruent condition, target objects were placed in background images that depicted a context in which the target object would not normally be found. In this condition, target objects were placed in an unnaturally fitting, but physically plausible area of the scene (e.g., a sign post in a backyard, bottom left corner of the image). An effort was made to place the target objects in similar spatial locations across the context conditions, however, the semantic structure of the scenes meant that physically plausible target locations often differed across conditions. Nonetheless, the mean target eccentricity from the center of the screen did not differ across the congruent (M = 7.73°) and incongruent (M = 8.35°) conditions, t (124) = .99, p = .32. See Figure 1 for an example of a congruent and incongruent background-plus-target image.

There has been some speculation that the line drawings used by Loftus and Mackworth (1978) may have differed across the semantic congruency conditions in terms of perceptual salience (Henderson & Hollingworth, 1999; Henderson et al., 1999; Underwood & Foulsham, 2006; Vo & Henderson, 2009; 2011); that is, perceptual rather

than semantic factors may have lead to the earlier fixations on incongruent targets they report. As a way to mitigate the contamination of perceptual salience on any differences measured across our semantic condition, the images used in the current experiment were submitted to a perceptual salience analysis (Zhang, Tong, Marks, Shan, & Cottrell, 2008). The salience analysis we chose uses Bayesian probability coupled with difference of Gaussians (DoG) filters on the intensity and color channels to compute local informativeness (i.e., rarity) at each pixel location. Higher salience scores indicate a greater degree of perceptual salience. Zhang et al. (2008) have shown that this salience algorithm performs as well or better than traditional salience analyses (Bruce & Tsotsos, 2006; Gao & Vasconcelos, 2007; Itti, Koch, & Nibur, 1998) at predicting human observer eye-movements. An example of salience feature maps applied to congruent and incongruent background-plus-target images is presented in Figure 2.

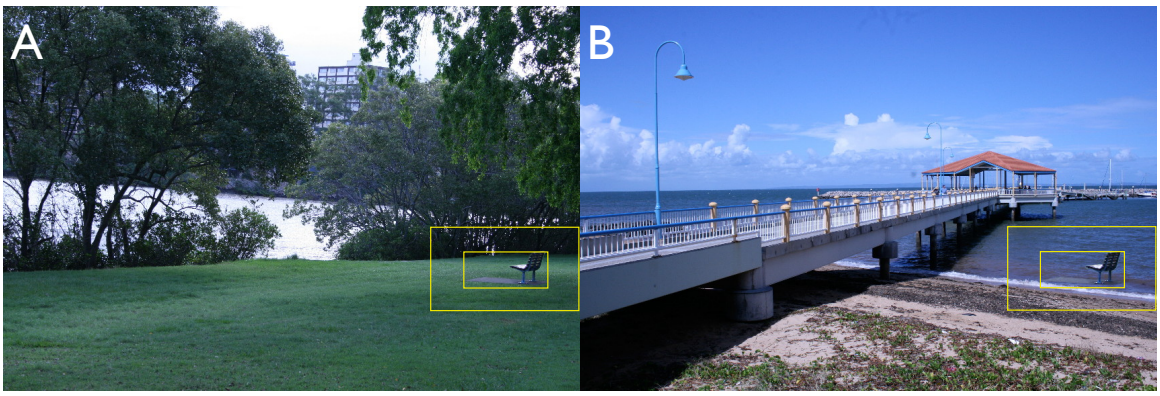

Figure 1. (A) An example of a background-plus-target image from the congruent condition, including the target area of interest and the peripheral target area of interest. (B) An example of a background-plus-target image from the incongruent condition, including the target area of interest and the peripheral target area of interest.

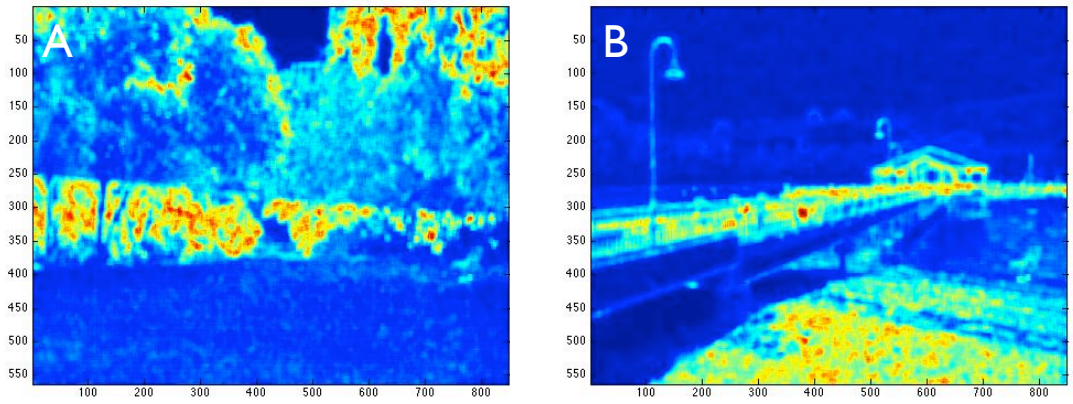

Figure 2. (A) An example of a congruent background-plus-target image with a salience map applied, where warmer colours indicate higher degrees of featural salience. (B) An example of an incongruent background-plus-target image with a salience map applied. The x- and y-axes indicate pixel location.

Unlike some other salience models (Bruce & Tsotsos, 2006; Gao & Vasconcelos, 2007; Itti et al., 1998; Torralba, Oliva, Castelhano, & Henderson, 2006) the algorithm provided by Zhang et al. (2008) does not compute a global salience statistic. As such, we calculated salience in the two following ways - target salience and target change salience. We defined target salience as the salience value in a rectangular area surrounding the target object in the background-plus-target images. To calculate the target salience values, we submitted each background-plus-target image to the salience analysis and calculated a salience score for each pixel in the target area, similar to the target area of interest used in the eye-movement measurements described below. Given the target area was matched across the context conditions, the sum of the salience values for each pixel in the target area provided a target salience score that was unbiased across the context conditions. Mean target salience values for the image set were then computed separately for congruent and

incongruent images. Mean target salience for the congruent (M = 135.67) and incongruent (M = 134.35) images did not differ significantly, t (122) = 0.15, p = 0.88.

For the target change salience, we submitted both the background-plus-target image and the background-only image to a salience analysis. We computed a salience value for the target area in both images (i.e., when the target was present and when the target was absent). We again computed salience values for each pixel in the target area for both images and then subtracted the target-absent salience value from the target-present salience value to provide a target salience change value for each pair of images. Mean target salience change values for the image set were then computed separately for congruent and incongruent images. Mean target salience change for the congruent (M = 11.59) and incongruent (M = 13.49) images did not differ significantly, t (122) = 0.25, p = 0.80.

Stimuli were presented on a 20" Viewsonic P220f monitor at a screen resolution of 1600 x 1200 pixels using Experiment Builder (SR Research, Canada). Participants sat 58 cm from the computer screen with their head stabilized using a chin rest. Eye-movements were measured using an EyeLink II head-mounted system (SR Research, Canada). Measurements from the right-eye only were recorded, however, binocular vision was used for scene exploration. Participants used a standard computer keyboard to make manual responses.

Procedure. After signing an experimental consent form, verbal task instructions were given to participants. The eye-gaze tracking headset was then put on the participant

and adjusted for comfort. The participant's eye-gaze was then calibrated and validated before beginning the experiment.

Each trial began with a calibration check. If the participant's eye-gaze needed to be re-calibrated, the experimenter adjusted the calibration. When no further calibration was needed, the participant pressed the spacebar to begin the trial. A fixation cross was presented centred on the computer screen for 500 ms, and was used to standardize eye-gaze prior to the presentation of the test stimuli. The image containing both the background and target object (*A*) was then presented for 250 ms, followed by a blank white screen for 250 ms. The background-only image (*A'*) was then presented for 250 ms, followed by another blank white screen for 250 ms. This sequence, from the first image (*A*) to the blank screen following the second image (*A'*) was presented repeatedly for up to 19 cycles or until a response was made. Participants were asked to press the spacebar when they knew which object was changing from *A* to *A'* and could accurately report its identity. When the spacebar was pressed, the image present at that time was replaced by a prompt asking participants to say aloud the identity of the target object using a single word. The experimenter recorded the vocal identification responses.

Each participant was presented with all 126 image pairs (63 congruent image pairs, 63 incongruent image pairs) in random order.

Results and Discussion

The purpose of the current experiment was to examine the processes that underlie performance benefits for semantically incongruent targets observed commonly in change

detection tasks. To that end, we report first behavioural measures to establish that change detection performance is indeed superior for incongruent targets, and then eye-movement measures to examine the nature of the processes that produce this benefit for incongruent targets.

Behavioural Measures

We used three behavioural measures to explore performance in the flicker task: mean response time (RT), misses, and errors. RTs were calculated for correctly responded to trials and denote the time from which the first image (*A'*) was presented to the time at which the button press was made on each trial. Misses were defined as trials in which no manual response was given presumably because the change was not detected. Errors were defined as trials in which a manual response was given, but an incorrect object label was reported. Mean RTs, and proportions of misses and errors were calculated separately for the congruent and incongruent conditions for each participant, and these data were then compared using paired-sample *t*-tests. A summary of the behavioural measures is presented in Table 1.

Table 1. A summary of the behavioural measures used in the current experiment for each context condition, including mean response time, mean proportion of changes missed, and mean proportion of object labelling errors, with standard error of within-subject variation in brackets (Morey, 2008).

	Semantic Congruity	
	Congruent	Incongruent
Response Time (ms)	4778 (140)	3761 (140)
Misses (prop.)	.09 (.01)	.05 (.01)
Errors (prop.)	.03 (.01)	.04 (.01)

Response Times. Participants responded significantly faster to target objects embedded within an incongruent context (M = 3761 ms) than to target objects embedded within a congruent context (M = 4778 ms), t (14) = 7.26, $p < .001$.

Misses. Participants missed significantly fewer changes on incongruent trials (M = .05) than on congruent trials (M = .09), t (14) = 3.56, p = .003.

Errors. There was no difference in the proportion of labeling errors across the two context conditions, t (14) = 1.10, p = .29.

Summary. The behavioural measures corroborate previous research showing a performance benefit for semantically incongruent objects using a change detection task (e.g., Hollingworth & Henderson, 2009; LaPointe et al., 2013). In this experiment, participants were more likely and faster to respond on trials containing a semantically incongruent target object than on trials containing a semantically congruent target object.

Eye Movement Measures

To examine the time course and nature of the performance benefits for semantically incongruent targets uncovered by our behavioural measures, we turn now to participants' eye movement behaviour. Eye movements were recorded for the entirety of each trial, from the presentation of the fixation cross until the trial expired or a button press was recorded. That is, eye movements were recorded during the presentation of both scenes (*A* & *A'*), as well as during the blank screens between the two scenes. In line with previous research (Vo & Henderson, 2009; 2011), we defined the target area as a rectangle just large enough to encompass the target object. Because the same target object appeared in both congruent and

incongruent conditions, the size of the target object areas were matched across conditions. In addition, we defined a peripheral target area to be a rectangle extending 6° of visual angle larger than the target area in both length and width. A summary of eye movement measures is presented in Table 2.

Table 2. A summary of the eye movement measures used in the current experiment. These measures were used to assess the degree to which eye gaze is attracted to, or fails to disengage from, the target objects, averaged across participants for each context condition, with standard error of within condition variation in brackets.

	Semantic Congruity	
	Congruent	Incongruent
Initial saccade latency (ms)	444 (10.79)	416 (10.79)
Initial saccade amplitude (deg.)	4.97 (.12)	5.23 (.12)
Initial saccade error (deg.)	91.34 (2.87)	84.40 (2.87)
Target first saccade amplitude (deg.)	4.53 (.29)	4.82 (.29)
Target last saccade amplitude (deg.)	2.71 (.46)	2.91 (.46)
Target first fixation latency (ms)	2491 (119.85)	1833 (119.85)
Prob. of fixating target peripherally - foveally	.17 (.01)	.23 (.01)
Target first fixation duration (ms)	496 (30.60)	515 (30.60)
Target last fixation duration (ms)	583 (32.48)	556 (32.48)
Target fixation dwell time (ms)	765 (75.74)	882 (75.74)

The eye movement measures we chose are a combination of those used previously in the literature and a number of new measures. Use of these measures was aimed at examining the nature of scene processing during the first few moments of visual perception of a natural scene, as well as the nature of processing once the eyes have landed on the target object. We have divided the measures into two categories: those for which the

attention attraction hypothesis makes a clear prediction and those for which the attention disengagement hypothesis makes a clear prediction.[1] [2] Again, it is important to note the attention attraction and attention disengagement hypotheses are not mutually exclusive. It is conceivable that attention is initially attracted to semantically incongruent objects and lingers on these informative objects, perhaps in an attempt to reconcile their identities with the surrounding context.

Attention Attraction

The attention attraction hypothesis predicts that attention is drawn to objects that are in semantic conflict with the context of the scene, and that this attraction should occur early during scene processing. For each of the following eye movement measures, the attention attraction hypothesis makes clear predictions of differences across the two context conditions. The attention disengagement hypothesis, on the other hand, predicts that attention is dispersed randomly about the scene until it has focused on a semantically incongruent object, at which point it lingers on these objects. For each of the following eye movement measures, the attention disengagement hypothesis predicts no difference across the two context conditions.

Initial Saccade Latency. Defined as the time elapsed prior to participants making their first saccade following the presentation of the first image (*A'*) for each trial, initial saccade latency has been used previously in the literature to investigate whether the semantic congruency of target objects differentially affects early eye movement patterns during scene viewing (Vo & Henderson, 2009; 2011). While this measure does not include

information about where the saccade is moving, shorter initial saccade latency for incongruent than congruent trials would be considered consistent with the attention attraction hypothesis, presumably because attention is being pulled by the incongruent objects early during scene viewing. Indeed, in the present study, participants were faster to make their initial saccade on incongruent trials (M = 416 ms) than on congruent trials (M = 444 ms), t (14) = 2.49, p = .03.

Initial Saccade Amplitude. Defined as the size of the initial saccade made following the presentation of the first image (*A'*) and measured in visual degrees, initial saccade amplitude has been used previously in the literature to investigate whether semantic congruency of target objects differentially affects early eye movement patterns during scene viewing (Henderson et al., 1999; Vo & Henderson, 2009; 2011). Similar to the initial saccade latency, initial saccade amplitude does not address the direction of the initial saccade. However, larger saccade amplitude for incongruent than congruent trials would appear to be consistent with the attention attraction hypothesis, perhaps as the draw of attention to incongruent objects contributes to larger initial saccades upon first viewing the scene. In the present study, the difference between initial saccade amplitude for incongruent target objects (M = 5.23°) and congruent target objects (M = 4.97°) approached significance, t (14) = 1.98, p = .07.

Initial Saccade Error. Whereas the two previous measures established the latency and size of the first saccade upon viewing the scene, initial saccade error was used to measure its direction. We calculated this measure as the difference between the angle

defined by the initial saccade and the angle defined by the location of the target object, both relative to the centrally located fixation cross. An initial saccade error of 0° would indicate the participants' eyes moved directly towards the target object, an error rate of 180° would indicate the eyes moved in the opposite direction of the target object, and an error rate of 90° would indicate eye movements based on chance. Smaller initial saccade errors for incongruent than congruent trials would be consistent with the attention attraction hypothesis. A related measure was used to study congruency effects on early stages of scene processing by Vo and Henderson (2009). However, their measure of saccade accuracy recorded only whether or not the saccade was directed to the same side of the screen as the target object. In the present study, initial saccade error was lower for incongruent targets (M = 84.4°) than for congruent targets (M = 91.34°), t (14) = 2.42, p = .03. Furthermore, initial saccade error for incongruent targets differed significantly from chance, t (14) = 1.76, p = .05, whereas initial saccade error for congruent targets did not, t (14) = .68, p = .25.

It is possible that the difference in initial saccade error measured across our context conditions is driven primarily by target objects located close to the fixation cross, rather than those located further away from initial fixation. To test this possibility, we compared the eccentricity of each target object from the fixation cross to the degree of initial saccade error for our incongruent context condition. If the eccentricity of the target object is driving the initial saccade error rate we should see a positive correlation between these measures. However, we failed to find a correlation between the eccentricity of the incongruent object

locations (M = 8.35°) and the initial saccade error rates on those trials (M = 84.4°), r = .20, p = .11 (see Figure 3).

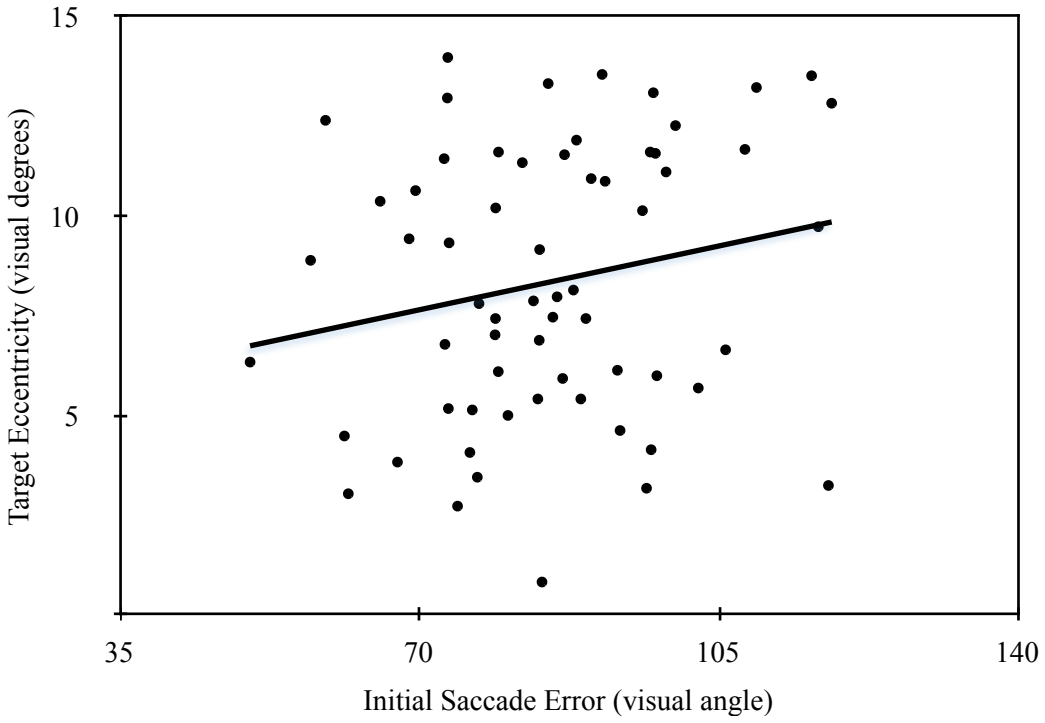

Figure 3. The correlation between incongruent targets' eccentricity (measured as the distance from the central fixation point in visual degrees) and initial saccade error rates (measured in visual angle) on the corresponding incongruent trials.

Target First Saccade Amplitude. Defined as the size of the first saccade to enter the target area and measured in degrees of visual angle, target first saccade amplitude has been used previously (Loftus & Mackworth, 1978; Vo & Henderson, 2009; 2011) as a measure of attention attraction to the target object. If attention is attracted to semantically incongruent objects, as proposed under the attention attraction hypothesis, rather than attention being deployed randomly, as proposed under the attention disengagement

hypothesis, we might expect to find larger first saccade amplitudes to the target object on incongruent trials relative to congruent trials. In the present study, however, we found no difference in the size of the saccade to first enter the target area, t (14) = 0.66, p = .52.

Target Last Saccade Amplitude. This measure is defined as the size of the last saccade to enter the target area prior to the end of the trial (either due to button press or trial expiration). Similar to the previous measure, if attention is attracted to semantically incongruent objects, rather than being dispersed randomly, we might expect larger final target saccades on incongruent trials relative to congruent trials. Although the previous measure may have captured influences of context congruency early during scene perception, the current measure may capture influences of context congruency just prior to the detection of the target object. In this case, however, we found no difference between the two context conditions, t (14) = 0.44, p = .67.

Target First Fixation Latency. Defined as the time taken to first fixate the target object, target first fixation latency has been used previously (Vo & Henderson, 2009; 2011) to investigate eye movement behaviour early during scene processing. If attention is attracted to incongruent objects, rather than being deployed randomly, we should expect incongruent targets to be first fixated earlier than congruent targets. Here, the difference between target first fixation latencies for incongruent targets (M = 1834 ms) and congruent targets (M = 2491 ms), approached significance, t (14) = 2.06, p = .06, with faster latencies for incongruent targets.

Probability of Fixating the Target Peripherally and then Foveally. This measure is defined as the likelihood of fixating the target area immediately following fixating the peripheral target area (which extended beyond the target area in both width and length by 6° of visual angle). This measure might well be larger for incongruent than congruent trials if fixating the peripheral target area increases the likelihood of having attention subsequently drawn to the target area, an idea generally consistent with the attention attraction hypothesis. In the present experiment, we found that participants were indeed more likely to fixate the target object directly after fixating the peripheral target area on incongruent trials (M = .23) than on congruent trials (M =.17), t (14) = 4.48, $p < .01$.

A strong variant of the attention attraction hypothesis predicts that attention ought to be attracted to a semantically incongruent target object to the same degree regardless of the distance between current fixation and the target object. However, a weaker variant of this hypothesis might predict that attention would be attracted to semantically incongruent objects only when the current fixation is relatively close to the target (i.e., within the near visual periphery). Furthermore, this weaker variant of the attention attraction hypothesis might predict that faster change detection for incongruent than for congruent targets would depend heavily on trials in which participants fixate the peripheral target area just prior to fixating the target area. To examine this issue, mean RTs for the two context conditions were computed separately for trials in which participants fixated the peripheral area just prior to fixating the target and trials in which this sequence of fixations did not occur. These mean RTs were submitted to an ANOVA that treated context (congruent/incongruent)

and fixation sequence (peripheral then target area/non-peripheral then target area) as within-subject variables. This analysis revealed a significant main effect of context, F (1, 13) = 19.58, $p < .001$, $\eta^2_p = .60$, with changes to incongruent targets (M = 4033 ms) being detected significantly faster than changes to congruent targets (M = 4827 ms). There was also a significant main effect of fixation sequence, F (1, 13) = 8.57, $p = .01$, $\eta^2_p = .40$, with changes being detected faster on trials in which the eyes fixated the target object immediately following the peripheral area (M = 4072 ms) compared with trials where no such sequence occurred (M = 4787 ms). Crucially, there was no interaction between these factors, F <1.

Probability of Early Target Fixation. We were interested in whether the likelihood of fixating incongruent targets differed from that for congruent targets for the first few fixations. The probability of early target fixation differences has been used previously in related studies of early scene processing (Loftus & Mackworth, 1978; Henderson et al., 1999). For this measure, the attention attraction hypothesis makes a strong prediction - if attention is attracted to incongruent objects, especially if this attraction occurs early during scene processing, the likelihood of fixating the incongruent targets should be higher than the likelihood of fixating congruent targets within the first few fixations of the scene. Following the lead of Henderson et al. (1999), we examined the probability of target fixation immediately after the first, second, third, and fourth fixations by participants. A summary of the cumulative probabilities of having fixated the target object as a function of ordinal fixation number and target object congruency is presented in

Figure 4.

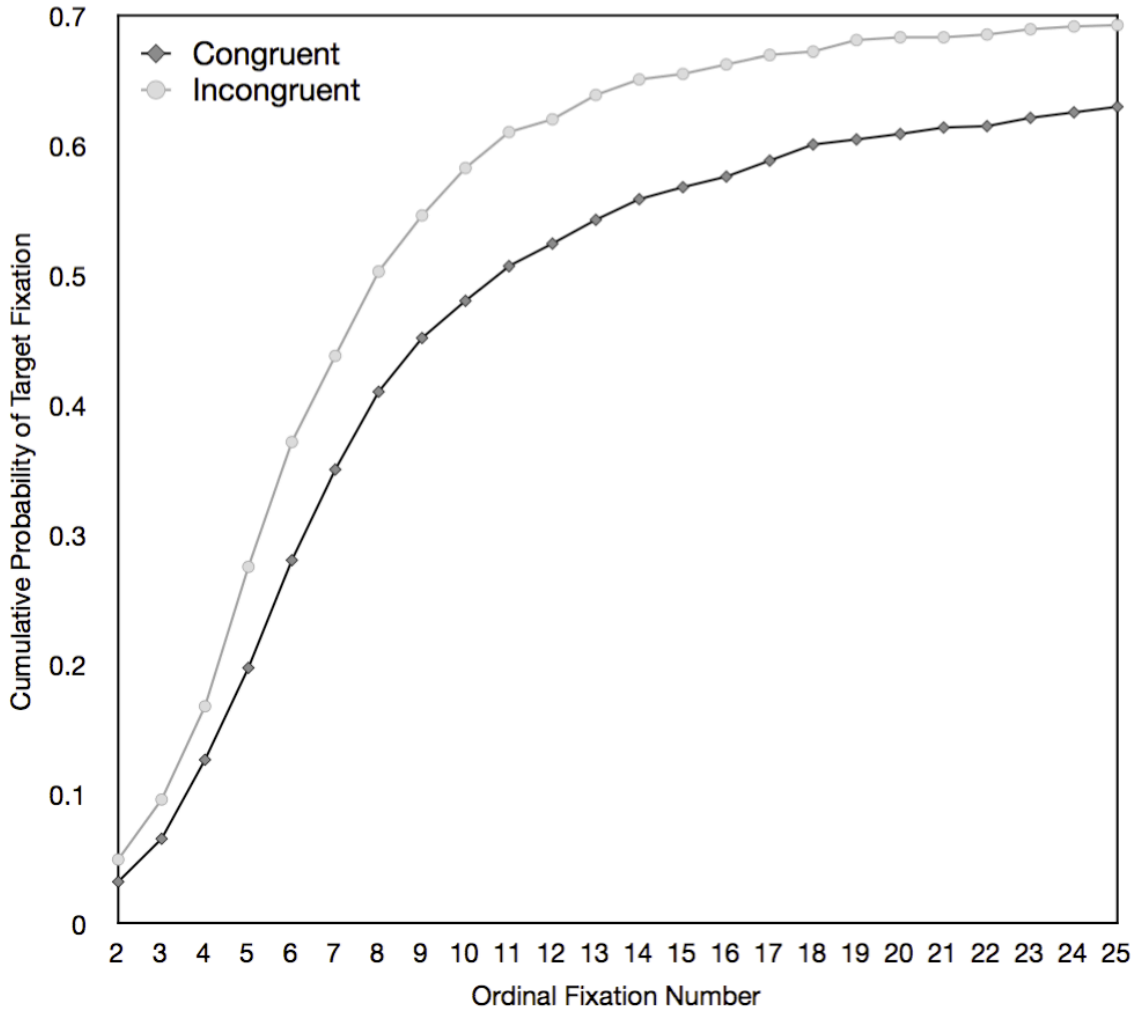

Figure 4. The cumulative probability of having fixated the target object as a function of the ordinal fixation number and semantic consistency. Note, on some trials the target object is never fixated.

The probability of fixating the target after the first saccade was .041. To evaluate whether this probability was affected by congruency, the probabilities of fixating congruent and incongruent targets after the first saccade were compared, and found not to differ

significantly (M = .03 for congruent, M = .05 for incongruent), t (14) = 1.43, p = .17. The probability of fixating the target after the second saccade was .08 and was also not influenced by context congruency, with the probability of fixating the congruent target at .07 and the probability of fixating the incongruent target at .10, t (14) = 1.63, p = .13. The probability of fixating the target after the third saccade was .15, again not differing significantly between congruent (M = .13) and incongruent (M = .17) conditions, t (14) = 1.93, p = .07. Finally, the probability of fixating the target after the fourth saccade was .24, and was significantly different across congruent (M = .20) and incongruent (M = .28) conditions, t (14) = 2.76, p = .02. The probability of fixating the target after the fifth saccade up until after the 25th saccade continued to differ significantly across the context condition with a bias in favour of fixating incongruent targets.

Summary. There appears to be strong evidence that eye-gaze is attracted towards semantically incongruent objects. Participants' initial saccade following image onset occurred earlier, was marginally larger, and was directed closer to incongruent targets than to congruent targets. Moreover, participants fixated the incongruent targets earlier than congruent targets. These findings complement others in support of the view that attention can be attracted to incongruent objects, not only preferentially, but early during scene processing (Becker, Pashler, & Lubin, 2007; Bonitz & Gordon, 2008; Gordon, 2004; 2006; Joubert, Rousselet, Fize, & Fabre-Thorpe, 2007; Loftus & Mackworth, 1978; Underwood & Foulsham, 2006; Underwood, Humphreys, & Cross, 2007; Underwood, Templeman, Lamming, & Foulsham, 2008).

Attention Disengagement

Despite finding strong evidence that attention is attracted to semantically incongruent objects, these results are not necessarily at odds with the attention disengagement hypothesis. According to this view, attention is deployed randomly about the scene. As attention lands on an object that is semantically incongruent (or, presumably, informative in any sense) it lingers, perhaps as information is being retrieved. Although we have shown that attention (or, at least, eye gaze) is not deployed randomly about the scene, but rather it is driven towards incongruent objects, it may still be the case that attention lingers on incongruent objects once the eyes have landed in that area of the scene. The following eye movement measurements were chosen specifically to assess whether eye gaze lingers preferentially on semantically incongruent objects.

Target First Fixation Duration. Defined as the length of time participants' eyes remained on the target object the first time it was fixated, target first fixation duration has been used extensively in prior related studies (Bonitz & Gordon, 2008; Friedman, 1979; Henderson et al., 1999; Vo & Henderson, 2009; 2011). If there is a tendency for attention to linger on semantically incongruent objects longer than congruent objects, we should find the duration of the first target fixation to be longer for incongruent objects. Here, however, we found no difference in the duration of the first fixation on the target across our two context conditions, t (14) = 0.64, p = 0.54.

Target Last Fixation Duration. Defined as the length of time participants' eyes remained on the target object the last time it was fixated prior to making a manual response

or the expiration of the trial, target last fixation duration has been used previously to study early scene processing (Friedman, 1979). Similar to the previous measure, longer last fixation durations for semantically incongruent targets than for congruent targets would be consistent with the idea that attention fails to disengage from incongruent objects. However, we found no difference in the duration of the last fixation between our two context conditions, t (14) = 0.83, p = 0.42.

Target Fixation Dwell Time. This measure is defined as the combined duration of all target fixations per trial, and has been used extensively in previous related studies (Bonitz & Gordon, 2008; Henderson et al., 1999; Vo & Henderson, 2009; 2011). If attention is more apt to linger on semantically incongruent objects, we should find longer overall fixation dwell times for these targets than for congruent targets. In the present experiment, we found no difference between the two context conditions, t (14) = 1.55, p = 0.14.

Summary. Using the above eye movement measures, we failed to find any evidence that attention lingers on objects that are in semantic conflict with the surrounding context. Given these results, we are led to conclude that performance benefits in this task are overwhelmingly driven by attention being attracted to semantically incongruent objects.

General Discussion

The behavioural and eye movement measures reported here are consistent with the view that semantically incongruent target objects attract overt attention in the context of a change detection task. As in prior studies, participants in the current study were faster and

more likely to detect changes to objects that were semantically incongruent with the context in which they were embedded (Bonitz & Gordon, 2008; Hollingworth & Henderson; 2000; 2007; Hollingworth, Williams, & Henderson, 2001; LaPointe et al., 2013). Furthermore, across the eye movement measures reported here, there was strong evidence that attention is attracted to semantically incongruent objects.

While viewing natural scenes for changes, participants' initial saccades were earlier, larger, and biased in direction toward the target object when that object was semantically incongruent with the context of the scene. Moreover, when participants' eyes landed within 6° of the target area, they were more likely to continue towards the incongruent target object. Finally, participants were more likely to fixate the incongruent target within the first four fixations of the scene. These findings suggest attention is attracted to objects that are semantically incongruent early during scene processing. In contrast, the eye movement measures offered little evidence that attention lingers on incongruent objects. There was no difference in the duration of the first, last, or overall fixation time across our context conditions.

One explanation for these findings is that the semantic qualities of the incongruent target objects draw attention, as posited under the attention attraction hypothesis. Specifically, as attention moves from the extraction of contextual information to local object information, the conflict in semantic information between the context and the incongruent object draws attention. According to this view, semantically incongruent objects are at least partially identified prior to attention being attracted to that area of the

scene. In this case, the conflict in meaning between the target object and scene context attracts attention for further processing. This theoretical account might be referred to as a strong variant of the attention attraction hypothesis. A strong variant of the attention attraction hypothesis is compelling in that it contradicts early selection views of attention. According to early selection theory, perceptual features of objects can be processed in parallel across the entirety of a visual scene, but serial shifts of focal attention are necessary to extract meaning associated with any particular object in a visual scene (Broadbent, 1958; for a review see Lachter, Forster, & Ruthruff, 2004). Attraction of attention to incongruent objects contradicts this view as it implies that access to meaning precedes the very shift of attention that is presumed by early section theories to provide access to that meaning. Given the strong implication for early selection theories of attention, it is worth considering in more detail whether the results reported here necessarily favour a strong variant of the attention attraction hypothesis.

In contrast to the strong variant of the attention attraction hypothesis, Vo and Henderson (2011) offered a weaker variant. These researchers describe visual scene processing as beginning in much the same way for scenes containing either a semantically congruent or incongruent target, with attention dispersed randomly about the scene in both cases. In this sense, extrafoveal processing of semantic information does not occur for either object type. However, as the eyes land close to an incongruent target, within the near visual periphery, attention is drawn to that object, thereby producing performance benefits relative to congruent targets. This variant of the attention attraction hypothesis conflicts

less with early selection views of attention in that it requires semantic information to be extracted near the fovea, but not necessarily in parallel across the entire visual scene.

In the current experiment, once the eyes had fixated the peripheral target area, the eyes were more likely to then fixate the target object if that object was semantically incongruent with the context of the scene. This finding appears to support the weaker version of the attention attraction hypothesis suggested by Vo and Henderson (2011). However, the RT benefit for incongruent targets over congruent targets was no larger on trials where the eyes landed in the peripheral target area prior to target fixation compared with trials where this fixation sequence was not observed. As such, the higher likelihood of fixating the target immediately following the peripheral area for incongruent objects might reflect a general extrafoveal attraction of attention towards semantic conflict.

A study reported by Gordon (2006), however, also urges caution before accepting the strong variant of the attention attraction hypothesis. In this study, participants were presented with a line drawing of a complex scene for 100 ms, followed by a pattern mask for 10 ms, a blank screen for 100 ms, and then a letter string. Participants were to indicate whether the letter string formed a legal English word (i.e., a lexical decision task). Crucially, the legal letter strings described either an object that was: (a) congruent with and present in the preceding scene; (b) congruent with but absent from the preceding scene; (c) incongruent with and present in the preceding scene; or (d) incongruent with but absent from the preceding scene. Gordon reasoned that if participants identified the incongruent targets during the short 100 ms presentation of the scene, then lexical decision times should

be facilitated for letter strings that matched those incongruent objects (i.e., positive priming). However, he found no priming at all for the incongruent letter strings. From this result, Gordon concluded that semantic processing of incongruent targets in parallel across a visual scene is likely not responsible for effects that appear to implicate attention attraction to those incongruent objects. Although one might worry that this priming method is simply insensitive to the processes that produce attention capture, another result in this study shows that this is not the case. Gordon observed negative priming for the congruent-present letter strings. Here, he reasoned attention was being drawn away from the congruent objects and towards the incongruent objects. Given the lack of priming for the incongruent letter strings, Gordon argued that the pull of attention towards these objects was not due to the semantic qualities of the incongruent objects, but rather to their perceptual qualities.

The results and interpretation offered by Gordon (2006) suggest yet another alternative account of the strong variant of the attention attraction hypothesis (Hollingworth & Henderson, 2000). As the viewer quickly recruits the semantic meaning of a scene, the perceptual characteristics of an incongruent object conflict with the perceptual expectations created by the scene context, and this conflict at the perceptual level attracts attention for further processing. Once attention is drawn to the perceptually disfluent object, attention lingers as the semantic meaning of the object is deciphered. The interpretation posited by Gordon does not necessarily violate the assumptions made under early selection models of attention if one assumes that a pre-attentive perceptual

expectation can be generated without identification of the incongruent object. Although perceptual salience of our target objects did not differ across our context conditions, Gordon's hypothesis assumes that it is not the perceptual qualities of the objects themselves, but rather the conflict between those qualities and the perceptual expectations set out by the context that attract attention.

In summary, the results of the current experiment suggest performance benefits for semantically incongruent objects in the flicker task can be explained by an attentional attraction to these objects, rather than a failure to disengage attention. The current results do not, however, distinguish whether this attraction is due to a semantic conflict or to perceptual disfluency involving the incongruent object and perceptual expectations driven by scene context. Further research is required to differentiate these two competing accounts. Moreover, although the current results suggest attention is attracted to semantically incongruent objects under conditions of viewing scenes for change, they do not address why the opposite effect often occurs under free viewing conditions. We might speculate that multiple processes operate when perceiving complex visual scenes, and the relative weighting of these processes depends on the specific task constraints. When viewing scenes for change under conditions of brief and interrupted scene presentation, as was the case in the current experiment, the viewer may be left especially sensitive to attention capture by mismatches between the scene context and target. In contrast, in tasks using extended and uninterrupted viewing, the viewer may be left especially sensitive to the structure inherent in a congruent context. Further research is required, perhaps in the

form of a detailed task analysis, to differentiate the processes underlying scene perception and their relative weightings when perceiving complex visual scenes under different task constraints.

Footnotes

Note 1. Target fixation count, defined as the mean number of fixations per trial, has been used previously in the literature to assess attention disengagement (Bonitz & Gordon, 2008; Henderson et al., 1999; Vo & Henderson, 2009), yet the attention attraction hypothesis would seem to make a similar prediction, albeit from a different mechanistic perspective. An extension of the attention disengagement hypothesis is that attention not only lingers on semantically incongruent objects, but also revisits these objects more often, perhaps in service of identifying the object. However, it is also possible that if attention is attracted to semantically incongruent objects, the object will be revisited until the change in noticed. As such, both hypotheses seem to make similar predictions. In any case, we failed to find any difference across the two context conditions, t (14) = 1.54, p = .15.

Note 2. We report no difference between the overall duration of target fixation time across our context conditions and interpret that result as lack of evidence for the attention disengagement hypothesis. We also report, however, that the mean length of the congruent trials was significantly longer than the incongruent trials. As such, in some sense, comparing total fixation time across the context conditions may not be a suitable contrast, given participants had more time to fixate the congruent trials. Accordingly, it may be argued the appropriate comparison is a proportionate measure of total fixation time to length of trial across the context conditions. In this case, a prediction can be made from the attention disengagement hypothesis – if attention lingers on semantically incongruent targets, we should find longer proportionate fixation dwell times on incongruent than congruent targets. It is likely, however, that a similar prediction could be made from the attention attraction hypothesis. As an example, assume the length of a congruent trial is 1000 ms and the length of an incongruent trial is 500 ms, with total target fixation durations of 250 ms for each. Further, assume that the 250 ms of total fixation duration is made up of five discrete target fixations of 50 ms each for both congruent and incongruent trials. Proportionally, the participant has spent more time fixating the incongruent targets. Yet, the total fixation time, the number of fixations, and the durations of those fixations are all equal. According to the attention attraction hypothesis, the larger proportion dwell time on incongruent trials is because attention is drawn to incongruent objects; whereas, the smaller proportion dwell time on congruent trials is because attention is dispersed about the scene randomly. As such, it is unclear how proportionate fixation duration differentiates the amount of time participants linger on one target type over the other. Regardless, we performed this comparison and found that participants spent proportionally more time fixating incongruent targets (M = .20) than congruent targets (M = .16), t (14) = 2.49, p = .03.

Chapter 4: Conflicting Effects of Context in Change Detection and Visual Search: A Dual Process Account

LaPointe, M. R. P., & Milliken, B. (in review)

Canadian Journal of Experimental Psychology,

Manuscript ID: CEP-2016-1316

Preface

The literature on complex scene perception has reported conflicting results both within and across tasks. In particular, some studies have reported context congruency benefits, whereas others have reported context incongruency benefits. In Chapter 1, it was demonstrated that multiple processes underlie complex scene perception and the relative weighting of these processes can be manipulated in such a way as to influence behaviour differentially within a task. It may be the case that the weighting of these processes are responsible for the conflicting results reported across studies and tasks as well. In particular, visual search tasks routinely produce congruency benefits, whereas change detection tasks routinely produce incongruency benefits. The purpose of Chapter 4 was to conduct a detailed task analysis by comparing performance in a change detection task to performance in a visual search task while keeping the stimulus set constant. In Experiment 1, characteristics of a typical visual search task were added to a change detection task. Here, the typical incongruency benefit was significantly reduced. In Experiment 2, a visual

search task was used to monitor the influence of context congruity on search performance. In this case, the incongruency benefit was reduced further. In Experiment 3, a hybrid change detection task was directly compared to a traditional visual search task. Across the three experiments, it was demonstrated that the relative weighting of the processes that underlie scene perception can be systematically manipulated in such a way as to reduce the performance benefits for contextually incongruent objects.

Abstract

Congruent contexts often facilitate performance in visual search and categorization tasks using natural scenes. A congruent context is thought to contain predictive information about the types of objects likely to be encountered, as well as their location. However, in change detection tasks, changes embedded in congruent contexts often produce impaired performance relative to incongruent contexts. Using a stimulus set controlled for object perceptual salience, we compare performance across change detection and visual search tasks, as well as a hybrid of these two tasks. The results support a dual process account with opposing influences of context congruency on change detection and object identification processes, which contribute differentially to performance in visual search and change detection tasks.

Introduction

Participants in psychology experiments are sometimes exposed to synthetic environments in which stimuli are presented in isolation. In contrast, in the natural world, flashes of light, bursts of sound, words, colours, and objects are usually, if not always, encountered within a broader context. Moreover, the role of context has played an important role in the study of a diverse range of topics in experimental psychology, including perception, attention, and memory. The present study focuses on the role of context in the perception of visual scenes.

An example of how context can impact the perception of visual scenes is provided by the phenomenon of contextual cueing (Chun & Jiang, 1998). Chun and Jiang had participants complete a standard visual search task with displays that included twelve letters, each letter appearing in one of 48 locations. Each search display contained a target letter T embedded among distractor letter L's, with each letter presented in one of four orientations. The goal of the participant was to identify the orientation of the target letter T as quickly as possible. The crucial manipulation was that 12 of the 24 trials in each of the 30 blocks contained a repeated spatial configuration of distractor objects from one block to the next. That is, the context defined by the spatial configuration of distractors was the same across all 30 blocks for targets on half of the trials. Participants produced search performance that was faster for targets that appeared in these consistent contexts, despite showing no explicit awareness of the repetitions of spatial configurations across the blocks.

This contextual cueing effect suggests that accrual of contextual instances can have a profound effect on visual behaviour.

Experiments using more naturalistic stimuli have also shown that context can shape the way we interact with objects. Palmer (1975) illustrated this point by presenting participants with a line drawing of a complex scene depicting a particular context (e.g., a kitchen), followed shortly after by the presentation of a line drawing of an isolated object (e.g., a kettle). Participants were asked to name the isolated object as quickly and accurately as possible. Interestingly, but perhaps not surprisingly, the objects that followed a context in which they would naturally fit were named faster and more accurately than those following an incongruent context. Palmer suggested that the context creates an expectation about what type of objects should be encountered, which in this case leads to faster naming times for congruent objects.

Similar benefits of a congruent context have been shown for rapid object categorization (Sun, Simon-Dack, Gordon, & Teder, 2011). Using a go/no-go task, Sun et al. presented participants with an image of an animal or vehicle for 20 ms on each trial. Following the presentation of the object, participants were required to make a manual response if the object had been an animal (go trials) or withhold a response if the object had been a vehicle (no-go trials). The crucial manipulation was that the objects were presented in isolation, embedded in a congruent context, or embedded in a phase-randomized scene background. Participants were faster to categorize the object in the congruent context condition than in the no-context and phase-randomized conditions.

Moreover, analysis of event-related potentials revealed that the latency of differential activity between go and no-go trials in frontal electrodes was slower by 20 ms in the phase-randomized condition than in the congruent context condition. In addition, larger frontal negativities and smaller late positive potentials were observed for the phase-randomized condition than for the congruent context condition. These results were interpreted as evidence that a congruent context facilitates object processing.

In line with the above results, many researchers have argued that visual scenes are processed first globally, during which gist information is coded, before moving towards more local processing of individual objects (Chun & Jiang, 1998; Navon, 1977; Schyns & Oliva, 1994). This claim has been substantiated by studies showing that gist information is processed at around 100 ms or less into scene processing (Biederman, 1981; Castelhano & Henderson, 2008; Fei-Fei, Iyer, Koch, & Perona, 2007; Henderson & Hollingworth, 1999; Intraub, 1981; Oliva & Schyns, 1997; Potter, 1975; 1976; Sampanes, Tseng, & Bridgeman, 2008; Schyns & Oliva, 1994), whereas local object information begins to be processed slightly later, at about 150 ms (VanRullen & Thorpe, 2001). As a viewer moves from global to local processing, a coherent context supports and facilitates the processing of congruent objects (Biederman, 1972; 1981; Biederman, Mezzanotte, & Rabinowitz, 1982; Boyce, Pollatsek & Rayner, 1989; Chun & Jiang, 1998; Friedman, 1979; Loftus & Mackworth, 1978; Mandler & Johnson, 1976; Palmer, 1975). The exploitation of a scene context appears to take advantage of the frequent co-occurrence of objects, the inherent structure of a congruent context, and prior experience with similar scenes. For example, prior

knowledge of a kitchen scene may lead attention towards the counter or stovetop in search of a kettle.

Empirical work has shown that prior semantic knowledge shapes the way in which we interact with complex scenes. For example, Neider and Zelinsky (2006) presented participants with a two-word description of a target object (e.g., BLUE JEEP) for 1 s, followed by a computer generated search display scene. Participants were to indicate whether the target object was present or absent in the search display as quickly and accurately as possible. The crucial manipulation was that the target could be an object that was semantically congruent or incongruent with the following scene. Moreover, within the congruent scenes, some areas of the scene were considered semantically appropriate for the target object, whereas other areas were not. Across two experiments, Neider and Zelinsky found participants were 17% and 21% faster at indicating the presence of the target object when the object was embedded in a congruent scene. Furthermore, by analyzing eye movements, the researchers noted a tendency for participants to move their eyes to contextually appropriate areas of the scene almost immediately following the presentation of the scene. These results indicate that prior knowledge of objects and the context in which they are usually encountered can have profound effects on the efficiency with which we interact with complex visual scenes.

The congruency of target objects to the backgrounds against which they are presented has also been shown to affect object identification and scene categorization. Davenport and Potter (2004) presented participants with an image of a complex scene for

80 ms, followed immediately by a mask. Following this sequence, participants were asked to identify the foreground object, the background, or both the foreground object and background, as accurately as possible. Across four experiments, participants were more accurate at identifying foreground objects when they appeared on congruent backgrounds, and backgrounds when they appeared paired with a congruent foreground object, than when the pair was incongruent. Davenport and Potter interpreted these findings as evidence that congruency of the object/background pair influences the identification of both, even when focusing on just one part of the pair. Moreover, the researchers argued that perception of the foreground object and background happens concurrently and interactively, suggesting the co-occurrence of object/background is exploited during natural scene processing.

The literature reviewed to this point indicates a strong reliance and exploitation of context and its structure, which facilitates the processing of congruent objects when navigating visual scenes. Such results suggest that prior knowledge of typical contexts, the semantic structure of those contexts, and the regular co-occurrence of objects is used to guide attention when viewing complex scenes. Yet, there is a growing number of studies in which the opposite, counter-intuitive effect, has been reported - performance benefits for semantically incongruent objects embedded within visual scenes (Bonitz & Gordon, 2008; Gordon, 2004; 2006; Hollingworth & Henderson, 2000; 2007; Hollingworth, Williams, & Henderson, 2001; LaPointe, Lupiáñez, & Milliken, 2013; LaPointe & Milliken, 2016).

Some of the demonstrations of incongruency benefits have been observed in change detection tasks. For example, Hollingworth and Henderson (2000) presented participants with a complex line drawing of a scene containing a target object that was either semantically congruent or incongruent with the context of the scene. Following the 250 ms scene presentation, the image was replaced by a white screen for 80 ms, followed then by the same scene either with or without the target object present. This sequence of scene presentations interleaved with white screens continued until participants were able to report whether or not a change had occurred. Although there was no difference in change detection accuracy across the context conditions, participants were significantly faster to respond on incongruent trials than congruent trials.

A performance benefit for semantically incongruent objects has also been reported in a relatively simple perceptual identification task. Gordon (2004) presented participants with a brief 147 ms exposure of a complex line drawing depicting a particular scene context. Each of the images contained either a semantically congruent or incongruent target object. The images were immediately replaced with a pattern mask containing a spatial probe depicting either a percent sign or an ampersand. Participants were tasked with determining the identity of the probe as quickly and accurately as possible. Crucially, the spatial probe appeared at either the same location as the target object in the preceding scene or in a location that did not contain an object in the preceding scene. In this case, participants were significantly faster at identifying the spatial probe when it appeared in the same location as the incongruent target object than when it appeared in the location of a

congruent target object. Gordon interpreted these findings to suggest that attention is drawn to semantically incongruent objects early during scene viewing.

In yet another task, participants were asked to visually inspect a series of complex line drawings depicting various scene contexts in preparation for a later scene recognition memory test (Loftus & Mackworth, 1978). Embedded within each of these images was a target object that was either semantically congruent or incongruent with the scene context. While exploring the scenes in advance of the memory test, participants' eye movements were directed towards the incongruent targets earlier and more often than they were to the congruent targets. Moreover, the saccades directed towards the incongruent targets were significantly larger than those directed towards the congruent targets. From these results, Loftus and Mackworth argued that eye gaze, and presumably attention, is drawn to objects that are semantically incongruent with the surrounding scene context.

The findings summarized above show that across several tasks there appears to be no clear and consistent set of empirical results that describe the relation between context congruency in visual scenes and efficiency of visual processing. The obvious and pressing question, then, is why do some studies of object processing within visual scenes produce performance benefits of context congruency, whereas others produce the opposite result. One difference that may explain the conflicting results across studies relates to the stimulus sets used, which varied from one study to the next across a number of dimensions, such as complexity and ecological validity. This criticism has been raised previously, specifically in response to the findings presented by Loftus and Mackworth (1978) and the sparse images

used in that study (Henderson, Weeks, & Hollingworth, 1999; Vo & Henderson, 1999). Another possible explanation for these competing findings focuses on task differences across the various studies. According to this view, tasks have a range of processing demands, and the effect of context congruency in any given task will depend on the relative weightings of all processes affected by context congruency in that particular task. Moreover, if some processes are facilitated by context congruency and others are impaired by context congruency, then differences in the relative weightings of these processes in performance could in principle produce opposite effects of context congruency across tasks.

A recent study offered preliminary support for this type of task analysis of context congruency effects in visual scene processing. Using a change detection task like that of Hollingworth and Henderson (2000), LaPointe et al. (2013) observed benefits for semantically incongruent objects. In a subsequent experiment, the blank white screens typically presented between images were removed, which rendered the detection of changes on the basis of luminance transients trivially easy. If the benefit for incongruent targets in the first experiment was caused by processes involved in more protracted detection of changes when luminance transients were not available, then this effect ought not to be observed here. Indeed, in this experiment performance was superior for congruent rather than for incongruent targets. The authors interpreted these findings as evidence that two processes are at work when viewing complex scenes: one process is related to

identification of objects and benefits from context/target congruency, whereas the other process is related to detection of objects and benefits from context/target incongruency.

The present study also focuses on a task analysis of context congruency effects in visual scene processing. The research strategy here was to examine the mapping between two tasks that have produced opposite effects of context congruency in prior studies: change detection and visual search. In particular, we start by measuring context incongruency benefits in a change detection task, and then measure context congruency effects in hybrid tasks that introduce components of more conventional visual search tasks. If different task demands on their own can produce opposite effects of context congruency on performance, then it ought to be possible to identify particular task properties that push context incongruency benefits observed previously in change detection in the direction of the context congruency benefits observed previously in visual search. To be clear, although prior studies have demonstrated opposite context congruency effects for these two tasks individually, no single study has demonstrated such effects with the stimulus set held constant.

Experiment 1

The purpose of Experiment 1 was two-fold: (1) to replicate previous findings showing a context congruency cost in a change detection task, and (2) to examine how this pattern of performance would be affected by introducing a key component of a visual search task to a change detection task. LaPointe et al. (2013) proposed that two processes, object detection and object identification, are affected in opposite ways by semantic

congruency between target objects and the scene context. Specifically, they proposed that object detection benefits from semantic incongruency between target and context, whereas object identification benefits from semantic congruency between target and context. Following on this proposal, our hypothesis was that change detection tasks depend heavily on object detection processes, whereas more conventional search tasks depend more on object identification processes. As such, modifying a change detection task by introducing components of more conventional search ought to shift context congruency effects away from context congruency costs and toward context congruency benefits.

One group in Experiment 1 completed a change detection task using natural scenes and similar task constraints to previous studies (Hollingworth & Henderson, 2000; LaPointe et al., 2013; Rensink, O'Regan, & Clark, 1997). The prediction for this group was that a processing emphasis on target detection would produce a performance benefit for context incongruency. For two other groups in Experiment 1, the change detection task was altered to introduce a key component of a visual search task. Specifically, one group of participants was presented with an image of the target object prior to the initiation of the alternating scenes, while another group was presented with a word describing the target object prior to the presentation of the alternating scenes. As such, both of these groups completed a task that may be thought of as a hybrid between change detection and visual search, as targets were defined both by target information presented to participants prior to the trial, and by the fact that the target was present in only one of the two alternating scenes. The introduction of target information prior to the onset of the alternating scenes

was aimed at increasing the emphasis on object identification relative to conventional change detection tasks. In particular, knowing what the target object is prior to the presentation of the scene ought to allow participants to take advantage of the structure of a congruent context to guide attention to areas of that scene that are likely to contain that target object. The prediction for these two groups, therefore, was that performance should shift toward a context congruency benefit. Indeed, in many studies that have reported context congruency benefits in scene perception, participants were given an object or category label prior to presentation of the visual scene (Henderson et al., 1999; Neider & Zelinsky, 2006; Torralba, Oliva, Castelhano, & Henderson, 2006; Vo & Henderson, 2011; Vo & Schneider, 2010).

To summarize, we hypothesized that for participants given no prior knowledge of the target object, the processing emphasis should be on object detection processes, and therefore change detection ought to be better for targets embedded in an incongruent context than for targets embedded in a congruent context (Bonitz & Gordon, 2008; Hollingworth & Henderson, 2000; 2007; Hollingworth, Williams, & Henderson, 2001; LaPointe et al., 2013; LaPointe & Milliken, 2016). In contrast, for participants given target information (i.e., image or word) prior to presentation of the alternating scene images, the processing emphasis should be shifted toward object identification processes, and therefore performance ought to shift in the direction of a performance benefit for congruent trials. The rationale for this prediction is that prior knowledge of the target object may be necessary for viewers to allocate attention to target-appropriate areas of the scene that are

naturally constrained by the contextual structure of the scene (Biederman, 1972; 1981; Biederman et al., 1982; Boyce et al., 1989; Davenport & Potter, 2004; Friedman, 1979; Mandler & Johnson, 1976; Neider & Zelinsky, 2006; Sun et al., 2011).

Method

Participants. Ninety undergraduate psychology students (63 female) from McMaster University ranging in age from 17 to 37 years (M = 19.02, SD = 3.00) volunteered to participate in exchange for partial course credit. All participants reported normal or corrected-to-normal vision.

Apparatus and Stimuli. The stimuli were presented on a 24-inch BENQ LCD monitor with a resolution of 1920 x 1080 pixels, using Livecode programming software. The images used in the current experiment were the same as those used by LaPointe and Milliken (2016) and were created from photographs taken in Brisbane, Australia. In each of the three between-subject conditions (no-prior, image-prior, word-prior), 126 image pairs of scenes were used, with half of those image pairs serving as background-only images (*A*) and the other half serving as background-plus-target images (*A'*). The background-only images were taken from photographs that depicted a variety of natural scenes that differed in a number of ways, including indoor and outdoor, lighting, complexity, and the number of objects. The background-plus-target images were created by digitally superimposing a target object, which had been taken from a separate photograph, onto a copy of a background-only image. Each target object was placed on both a contextually congruent and contextually incongruent background image. The target objects in each context

condition were placed in physically plausible locations, regardless of their contextual congruency and an effort was made to place the targets in a similar spatial location across the context conditions (e.g., a bear in a congruent forest, bottom left corner the image, a bear in an incongruent kitchen, bottom left corner of the image) All told, 63 image pairs (i.e., *A* and *A'*) were presented in the contextually congruent condition and 63 image pairs were presented in the contextually incongruent condition.

The perceptual salience of the target objects did not differ across the context conditions. Using an algorithm that relies on Bayesian probability and independent component analyses to compute the degree of rarity of each pixel in the image (Zhang, Tong, Marks, Shan, & Cottrell, 2008), LaPointe and Milliken (2016) calculated perceptual salience for each target object in two separate ways - target salience and target change salience. Target salience was calculated by summing the salience values for each pixel within a rectangular area just large enough to encompass the target object. Given the target objects were matched across the context conditions, the sizes of the rectangular areas were necessarily matched across the context conditions as well. The summed target salience values for each context condition were then compared and found not to be significantly different from one another. Target change salience was computed by taking the summed salience value from within the target area and subtracting the summed salience value from the same area on the background-only image (i.e., without the target object present). These values were then compared across the context conditions and found not to differ significantly from one another (for detailed analyses, see LaPointe & Milliken, 2016).

Design. One third of participants received no information about the target object prior to the presentation of the first image in the change detection task. Another third of participants were presented with an image of the target object at the beginning of each trial, which we call the image-prior condition. The images of the target objects were created in a similar manner to the background-plus-target images. That is, the target objects were digitally removed from photographs and presented as isolated objects. The target objects presented at the beginning of the trial appeared the same size as they would later appear in the background-plus-target images. The final third of participants were presented with a one-word description of the target object at the beginning of each trial, which we call the word-prior condition. In each of these three conditions, half of the scene images contained a target object that was congruent with the scene context and the other half contained a target object that was incongruent with scene context. Target prior knowledge (no-prior/ image-prior/word-prior) served as a between-subjects variable and context congruency (congruent/incongruent) served as a within-subject variable. All 126 image pairs (63 congruent image pairs, 63 incongruent image pairs) were presented to each participant in a random order.

Procedure. Upon signing an experimental consent form, participants were given task instructions verbally, they watched a video reiterating those instructions with a short demonstration of the task, and they were given an opportunity to ask clarifying questions. For the participants in the image-prior and word-prior conditions, each trial began with either an image of the target object or a one-word description of the target object,

respectively, presented for 2000 ms, followed by a fixation cross centred on the screen for 500 ms. Next, the background-plus-target object image appeared for 250 ms, followed by a white inter-stimulus interval (ISI) screen for 250 ms. The background-only image was then presented for 250 ms, followed by another ISI for 250 ms. The sequence from the background-plus-target image to the last ISI repeated for up to 19 cycles or until a button press was registered. Participants in the no-prior condition were presented with the same sequence of stimuli from the fixation cross onwards. All participants were asked to press the spacebar if and when they could determine which object was changing from one image presentation to the next. Upon pressing the spacebar, the stimulus present on screen was replaced with a nine-box grid. Participants were instructed to indicate the location of the changing object by pressing a spatially corresponding key on the number pad of the keyboard.

Results

Three dependent variables were analyzed in this experiment. Mean response times (RT) were calculated from the onset of the first image (*A'*) until the spacebar was pressed for correctly responded to trials only. Misses were defined as trials in which the spacebar was not pressed, and presumably the change was not detected. Errors were initially defined as trials in which the spacebar was pressed, but the incorrect target location was given. However, error rates using this method were higher than desired (> .15), and likely occurred on trials in which the changes were detected but localized somewhat imprecisely. Therefore, for this and subsequent experiments, we defined errors as trials in which the

spacebar was pressed, but a response other than the correct or adjacent target locations was given. This method of categorizing errors reduced the error rates substantially, and likely offers a better estimate of the frequency with which participants provided a localization response that did not correspond to the target object. RTs, proportions of misses, and proportions of errors were submitted to mixed factor analyses of variance that treated context congruency (congruent/incongruent) as a within-subject factor and target prior knowledge (no-prior/image-prior/word-prior) as a between-subjects factor. The mean RTs for each condition are displayed in Figure 1.

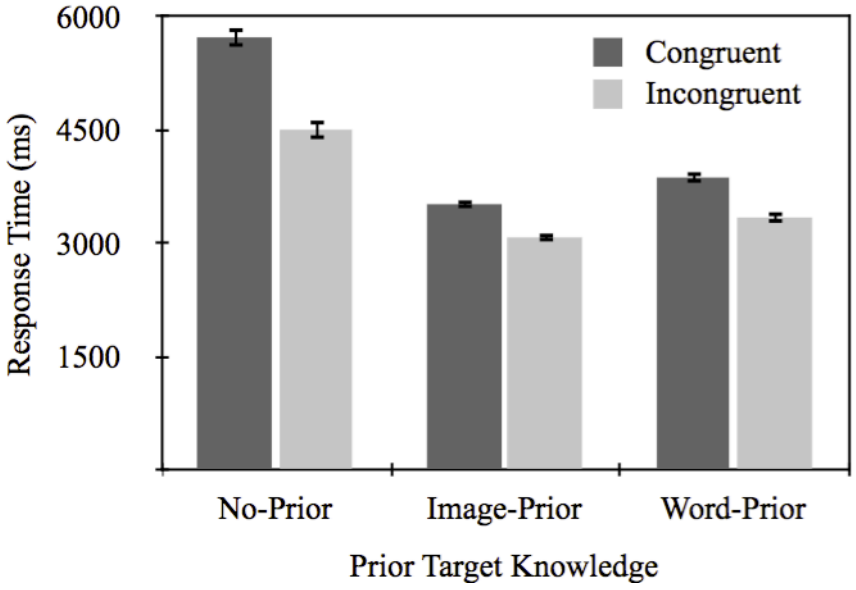

Figure 1. Mean response times (ms) for correctly detected changes in each context condition (congruent vs. incongruent) for each of the prior knowledge conditions (no-prior, image-prior, word-prior) in Experiment 1. Response times were measured from the onset of the first image (background-plus-target) until a keypress was recorded. Error bars indicate standard error of the mean corrected to remove between-subject variation (Morey, 2008).

Response Time. There was a significant main effect of context congruency, F (1, 87) = 201.05, $p < .001$, $\eta^2_p = .70$, with response time being slower on congruent trials (M = 4368 ms) than incongruent trials (M = 3638 ms). There was also a significant main effect of target prior knowledge, F (2, 87) = 9.37, $p < .001$, $\eta^2_p = .18$. A Fisher's LSD test that compared RTs across conditions revealed that participants were significantly slower to respond in the no-prior condition (M = 5111 ms) than in both the image-prior (M = 3295 ms) and word-prior conditions (M = 3603 ms), and the latter two conditions did not differ from one another ($LSD_{.05}$ = 1296 ms). Finally, there was a significant interaction between context congruency and target prior knowledge, F (2, 87) = 23.03, $p < .001$, $\eta^2_p = .35$. This interaction was examined further by analyzing the effect of context congruency separately for each target prior knowledge condition. Participants were significantly faster on incongruent trials (M = 4500 ms) than on congruent trials for all three target prior knowledge conditions (4500 ms vs 5721 ms for the no-prior condition, t (29) = 9.69, $p < .001$; 3076 ms vs 3515 ms for the image-prior condition, t (29) = 8.06, $p < .001$, and 3338 ms vs 3868 ms for the word-prior condition, t (29) = 7.50, $p < .001$). A Fisher's LSD test that compared the congruency effects across each pair of target prior knowledge conditions revealed that the congruency effect was larger for the no-prior condition than for both the image-prior and word-prior conditions, and that these latter two conditions did not differ from each other ($LSD_{.05}$ = 257 ms).

Errors. There were no significant effects in this analysis, $F < 1$ in all cases.

Misses. There was a significant main effect of context congruency, F (1, 87) = 25.62, $p < .001$, $\eta^2_p = .23$, with participants missing more changes on congruent trials (M = .043) than on incongruent trials (M = .026). There was also a significant main effect of target prior knowledge, F (2, 87) = 27.67, $p < .001$, $\eta^2_p = .39$. A Fisher's LSD test that compared the miss rates between the three target prior knowledge conditions revealed that more misses occurred for the no-prior condition (M = .06) than for both the image-prior (M = .02) and word-prior (M = .02) conditions, and that these latter two conditions did not differ from each other ($LSD_{.05}$ = .02). The interaction between context congruency and target prior knowledge was also significant, F (2, 87) = 11.98, $p < .001$, $\eta^2_p = .22$. This interaction was examined further by analyzing the effect of context congruency separately for each target prior knowledge condition. Participants missed significantly fewer changes on incongruent trials than on congruent trials for both the no-prior condition (.04 vs .08), t (29) = 4.98, $p < .001$, and the word-prior condition (.02 vs .03), t (29) = 2.24, $p = .03$. In contrast, for the image-prior condition, there was no difference in the proportion of missed changes between incongruent (M = .02) and congruent (M = .02) trials, t (29) = .56, $p = .58$.

Discussion

The results from the no-prior target condition replicate previous findings from studies using change detection tasks (Bonitz & Gordon, 2008; Hollingworth & Henderson, 2000; 2007; Hollingworth, Williams, & Henderson, 2001; LaPointe et al., 2013; LaPointe & Milliken, 2016). Participants were significantly faster and missed fewer changes when the target object was semantically incongruent than when the target object was semantically

congruent with the context of the scene. In the conditions that presented participants with target information prior to the alternating scenes (i.e., image-prior and word-prior conditions), participants produced a similar qualitative effect. In both of these conditions, participants were again significantly faster for semantically incongruent targets than semantically congruent targets. However, the incongruency effect was substantially smaller for these two conditions (439 ms for image-prior, 530 ms for word-prior) than for the no-prior condition (1271 ms).

By introducing a key property of visual search tasks (i.e., knowledge of the target prior to presentation of the scene) we assumed that processing demands would shift from an emphasis on object detection to an emphasis on object identification. For example, knowing the target object is a kettle should constrain attention to specific areas of a congruent kitchen context, but not to any particular area of an incongruent forest context. Indeed, the significantly smaller incongruency benefit observed for the image-prior and word-prior conditions than for the no-prior condition is consistent with the idea that a process that benefits from semantic congruency works in opposition to a process that benefits from semantic incongruency.

Experiment 2

The reduction of the incongruency benefit in Experiment 1 for participants who were given information about the target object before the change detection task suggests that this change in the task increased the contribution to performance of a process that benefits from semantic congruency between the target and scene context. At the same time,

it was notable that performance remained more efficient for incongruent than congruent trials. One possible account of this result is that when targets are defined both by information given prior to the alternating scenes, and by a change across the alternating scenes, participants may use either source of target information to guide performance. If this was the case for the image-prior and word-prior conditions in Experiment 1, then performance in these conditions might reflect a mixture of trials in which semantic congruency speeds performance and trials in which semantic congruency slows performance. The resulting mixture could well produce the results that were observed, in particular the semantic incongruency benefit may have happened to be larger, or occurred on more trials, than the semantic congruency benefit.

The purpose of the present experiment was to examine whether performance could be biased further in the direction of semantic congruency benefits by eliminating entirely the change detection component of the task. Specifically, rather than using a hybrid change detection/visual search task, we used a more conventional visual search task. Participants were given extended and uninterrupted viewing of the scene, and were required to search for a particular target that did not change in status across the duration of the trial. If change detection task properties (e.g., defining the target by change, or brief and rapid presentation of scenes) were responsible for biasing performance in favour of semantic incongruency benefits in Experiment 1, then we might expect to find semantic congruency benefits with the more conventional visual search task used here (Henderson et al., 1999; Neider & Zelinsky, 2006).

Method

Participants. Twenty-five undergraduates (21 female) from McMaster University ranging in age from 17 to 20 years (M = 18.36, SD = .70) volunteered to participate in exchange for partial course credit. Each participant self-reported normal or corrected-to-normal vision and none had participated in the previous experiment.

Apparatus and Stimuli. The background-plus-target images used in Experiment 1 were used in the current experiment, with 63 serving as congruent images and 63 serving as incongruent images. As in Experiment 1, the target objects did not differ in terms of perceptual salience across the context conditions. The 126 target-only images used in the image-prior condition in Experiment 1 were also used in the present experiment.

The apparatus used in Experiment 1 was also used in the present experiment.

Procedure. After signing an experimental consent form, participants were given task instructions verbally before watching a video tutorial that repeated the instructions and demonstrated an example of an experimental trial. Each trial began with the presentation of an image of an object for 2000 ms, followed by a fixation cross located in the centre of the screen for 500 ms. Participants were then presented an image of a natural scene that remained on the screen until the target object could be located and a manual response was made. Upon pressing the spacebar, the scene was replaced by a 9-box grid numbered 1-9. Participants were to indicate in which box the target object was located in the preceding scene by pressing the corresponding number on the keyboard. The object presented at the beginning of the trial was always the target object in the following scene. On half the trials

the target object and background scene were congruent, while on the other half of trials the object and background scene were incongruent. All 126 background-plus-target images (63 congruent, 63 incongruent) were presented to each participant in a random order.

Results

Two key dependent variables were analyzed in this experiment. As in Experiment 1, mean RTs were computed using correctly responded to trials only. Errors were defined as trials in which a location other than the target or adjacent locations was given. We did not analyze misses in this experiment because a response was given on every trial for all participants. Mean RTs and error rates were computed for each context condition separately for each participant and these data were submitted to two-tailed *t* tests for each measure.

Response Time. Participants responded significantly faster on incongruent trials (M = 1882 ms) than on congruent trials (M = 1940 ms), t (24) = 2.30, p = 0.03.

Errors. Participants made significantly fewer localization errors on incongruent trials (M = .01) than on congruent trials (M = .02), t (24) = 3.72, p < .001.

Discussion

In the present experiment, participants were presented with uninterrupted exposures of the scene images, in contrast to the brief and repeated scene exposures used in Experiment 1. Moreover, participants in the present experiment engaged in a visual search task, rather than change detection or hybrid change detection/visual search. Despite these changes participants continued to exhibit better performance on incongruent trials than congruent trials. It should be noted, however, that the response time benefit for incongruent

targets observed here (58 ms) was substantially smaller than the difference observed in the image-prior (439 ms), and word-prior (530 ms) conditions in Experiment 1, suggesting a further shift in weighting from object detection processes that benefit from semantic incongruency to object identification processes that benefit from semantic congruency. Nonetheless, the persistence of a significant incongruency benefit in the visual search task used in the present experiment remained a puzzle.

In re-examining the stimulus set used here, we identified a property that could well undermine performance on congruent trials specifically in tasks that require search for a known target. We noted that in some of the scene images there were multiple objects that belonged to the same semantic category as the target. For example, if the target object was a chair, there may have been more than one chair in the image. The additional object tokens were never an exact copy of the target object; they differed from the target object along a number of dimensions (e.g., size, color, orientation). However, they were undoubtedly tokens that represented the same object type. Moreover, the images that contained multiple tokens of the target object naturally occurred more often in congruent images than incongruent images - a target object chair was more likely to appear in a scene near other chairs in a congruent kitchen context than in an incongruent forest context. Further, when presented an image of the target object prior to the presentation of the scene, participants may then have found themselves comparing the target representation to each of the multiple tokens of that object in the search display to be certain which of those multiple similar tokens was actually the target. In retrospect, this issue seems likely to have slowed

response times and increased localization errors on congruent trials relative to incongruent trials here, and may well have done so for the image-prior and word-prior conditions in Experiment 1.

To examine whether the presence of multiple tokens of the target object contributed to the slower performance for congruent than incongruent trials, we re-analyzed both RTs and errors after removing all trials in which the image contained more than one token of the target object. Furthermore, to ensure that the same target object appeared in both a congruent and incongruent context, if an image was removed from the congruent context condition on the basis of multiple tokens of the target, we also removed the image containing the same target object from the incongruent context condition regardless of whether that image contained multiple tokens of the target. This stimulus selection process allowed us to compute mean RTs and error rates for a total of 64 trials (32 congruent, 32 incongruent) for each participant. Mean RTs for the two context conditions were not significantly different, with the numerical trend now favouring congruent trials (M = 1883 ms for congruent, M = 1902 ms for incongruent), $p > .10$. Error rates did not differ for congruent and incongruent items (M = .02 for both conditions).

In summary, when analyzing only trials in which the images did not contain multiple tokens of the target, there was no significant difference between performance on congruent and incongruent trials. This post-hoc analysis is consistent with the idea that the presence of multiple tokens of the target did slow search on congruent trials in the overall analysis. All told, the results from Experiments 1 and 2 suggest that multiple processes

underlie congruency effects in scene processing, and that visual search does introduce a shift in emphasis toward processes that benefit from congruency between the search target and scene context.

Experiment 3

The results of Experiment 1 demonstrated that a hybrid procedure in which search target information is provided prior to the change detection trial reduced the benefit for incongruent over congruent trials relative to a conventional change detection task. The results of Experiment 2 suggest that a conventional search task, in which all components of change detection are eliminated, reduces that incongruency benefit further. However, two limitations of Experiment 2 are noteworthy. First, search performance in Experiment 2 appears to have been affected by the multiple token issue described above. One purpose of Experiment 3 was to address this issue. Second, Experiment 2 did not provide a direct comparison of the hybrid and conventional search procedures. A second purpose of the present experiment was to make this direct comparison.

As such, two groups of participants completed Experiment 3. One group completed a conventional visual search task. The other group completed the hybrid change detection/ visual search task from Experiment 1, in which a target image was presented prior to a task that had alternating scenes with and without the target object. Importantly, all images that contained distractor objects that were similar in perceptual appearance to the target object were removed from the stimulus set. However, two important qualities of the stimulus set were retained. First, the same target object was presented in both a congruent and an

incongruent scene. This constraint was achieved by removing not only the images containing multiple tokens of the target object, but also the companion image from the other context condition (i.e., if a congruent chair image was removed, so too was the incongruent chair image). Second, the perceptual salience of the targets was re-analyzed to ensure that it was kept constant across the context conditions.

Method

Participants. Forty undergraduate students (32 female) from McMaster University ranging in age from 17 to 21 years (M = 18.33, SD = .73) participated in exchange for partial course credit. All participants reported normal or corrected-to-normal vision and none had participated in either of the first two experiments.

Apparatus and Stimuli. The images used in the current experiment were a subset of the images used in Experiments 1 and 2. After excluding images that contained distractor items that closely resembled the target object, the final stimulus set included sixty-four background-plus-target images, sixty-four background-only images, and thirty-two target images. Of the sixty-four background-plus-target and background-only images, half were included in the congruent condition and half were included in the incongruent condition. The target objects embedded in the scene images appeared in both a congruent context scene and an incongruent context scene.

Procedure. After signing an experimental consent form, participants were given task instructions verbally and then shown a video demonstrating the task. Half of the

participants participated in the hybrid change detection/visual search task and the other half participated in the conventional visual search task.

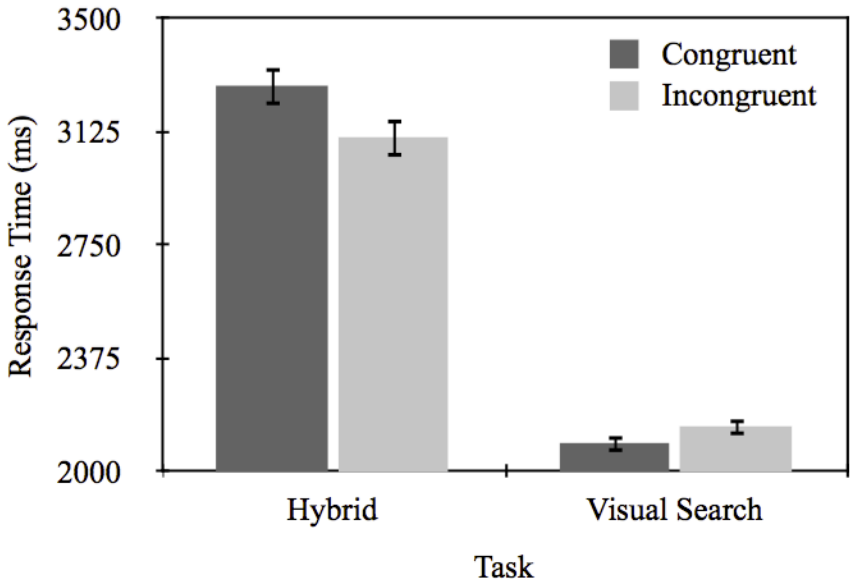

Figure 2. Mean response times (ms) for correctly responded to trials in each context condition (congruent vs. incongruent) for each task condition (hybrid vs. visual search) in Experiment 3. Response times were measured from the onset of the first image (background-plus-target) until a keypress was recorded. Error bars indicate standard error of the mean corrected to remove between-subject variation (Morey, 2008).

For those who participated in the hybrid task, each trial began with the presentation of an image of the target object for 2000 ms, followed by a fixation cross centered on screen for 500 ms. The background-plus-target object image was then presented for 250 ms, followed by a white ISI for 250 ms. The background-only image was then presented for 250 ms, followed by another white ISI for 250 ms. This sequence of displays repeated

until participants responded or up to a maximum of 19 cycles. Participants were asked to press the spacebar when they had localized the target object that was changing from one image presentation to the next. Note that the changing object was always the target object presented prior to the trial.

For those who participated in the conventional visual search task, each trial began with the presentation of an image of the target object for 2000 ms, followed by a fixation cross in the centre of the screen for 500 ms, followed by the presentation of the background-plus-target object image, which remained on the screen until a spacebar response was made. Participants were asked to press the spacebar once they had localized the target object.

In both task conditions, once the spacebar had been pressed the current image was removed from the screen and replaced by a grid the same size as the image, numbered 1 to 9. Using the number pad on the keyboard, participants were to indicate which box on the grid corresponded to the location of the target object in the preceding scene.

Results

Three dependent variables were used to assess behaviour in this experiment. Mean response times (RT) were calculated from the onset of the first image (*A'*) until the spacebar was pressed for correctly responded to trials only, and are displayed in Figure 2. Errors were defined as trials in which the spacebar was pressed, but neither the correct or adjacent locations were given. These measures were used to assess behaviour in the both the hybrid and visual search tasks. Misses were calculated as trials in which the spacebar

was not pressed. This measure was used to assess behaviour in the hybrid task only. RTs and proportions of errors were submitted to mixed factor analyses of variance that treated context congruency (congruent/incongruent) as a within-subject factor and task (hybrid/ visual search) as a between-subject factor. The proportions of misses were computed for each context congruency condition separately for each participant in the hybrid task and these data were submitted to a two-tailed *t* test.

Response Time. The main effect of context congruency did not reach significance, F (1, 38) = 2.62, p = .11, η^2_p = .06, but there was a significant main effect of task, F (1, 38) = 49.09, p < .001, η^2_p = .56. Participants responded faster in the visual search task (M = 2116 ms) than in the hybrid task (M = 3191 ms). Most important, there was a significant interaction between context congruency and task, F (1, 38) = 9.95, p = .003, η^2_p = .21. This interaction was examined further by analyzing the effect of context congruency separately for each task. For the hybrid task, participants responded significantly faster on incongruent trials (M = 3106 ms) than congruent trials (M = 3276 ms), t (19) = 2.63, p = .02. For the visual search task, the effect of context congruency approached significance, with participants responding faster on congruent trials (M = 2089 ms) than incongruent trials (M = 2143 ms), t (19) = 1.83, p = .08.

Errors. There were no significant effects in this analysis (all p > .10).

Misses. Participants responded on all trials in the visual search task, so analysis was conducted only for miss rates in the hybrid task. Participants missed significantly more

changes on congruent trials (M = .01) than incongruent trials (M = .003), t (19) = 2.03, p = .03.

Discussion

The results of this experiment demonstrate that semantic congruency between the target and scene context impacted performance differently in the hybrid and visual search tasks. Whereas performance was significantly better for incongruent than congruent trials in the hybrid task, the opposite trend was observed in the visual search task. These results are broadly consistent with the view that multiple processes contribute to context congruency effects in visual scene processing. More specifically, the results imply that processes that benefit from context congruency play a significantly greater role in the visual search task than in the hybrid change detection/visual search task, and that processes that benefit from context incongruency are predominant in the hybrid change detection/ visual search task. According to the framework introduced by LaPointe et al. (2013), we conclude that context congruency may have benefited target identification but hurt target detection, and that the visual search and hybrid tasks have significantly different weightings on these two processes. Specifically, target identification processes that constrain where a congruent target ought to appear may play a large role in the visual search task, whereas target detection processes that shift attention to incongruent targets quicker than congruent targets may play a large role in the hybrid task.

General Discussion

The purpose of the present study was to conduct a task analysis of context congruency effects in visual scene processing. This task analysis was prompted by the finding that context congruency sometimes benefits and other times hurts performance in studies of visual scene perception. The contrasting results could conceivably stem from stimulus set differences across studies. Alternatively, task differences alone might be responsible for shifts in context congruency effects across studies. To date, no study has carefully examined task effects while keeping the characteristics of the stimulus set constant.

In Experiment 1, a conventional change detection task produced a significant context incongruency benefit, replicating previous research (Hollingworth & Henderson, 2000; 2007; Hollingworth et al., 2001; LaPointe et al., 2013; LaPointe & Milliken, 2016). In a hybrid task, the incongruency benefit was significantly reduced. In Experiment 2, use of a conventional visual search task appeared to reduce the incongruency benefit further. However, a stimulus set issue in Experiment 2 also appeared to impact performance. Specifically, multiple tokens that corresponded to the target object type may have slowed performance for congruent relative to incongruent trials. In Experiment 3, the stimulus set was controlled for multiple tokens of the target object and a hybrid task was directly compared to a conventional visual search task. In this experiment, the context congruency effect differed qualitatively across tasks, with the hybrid task producing a significant incongruency benefit, and the conventional search task producing a congruency benefit that approached significance. Across the experiments, there was substantial evidence that

context congruency effects can vary markedly as a function of task when the stimulus set is held constant.

The results across three experiments suggest that multiple processes impact context congruency effects in the viewing of complex scenes, and that task demands influence the relative weighting of these processes. LaPointe et al. (2013) speculated that two independent processes underlie performance in studies of visual scene perception: object detection and object identification. On the one hand, early recruitment of a scene's gist information promotes expectancies regarding what types of objects ought to be encountered, as well as their likely locations within a scene. This process ought to benefit performance for trials in which object and context are semantically congruent, as semantic expectancies produce biases that correspond to the actual target. On the other hand, object detection may benefit from semantic incongruency between object and context, as attention may shift preferentially to areas in a scene that require further processing to resolve context-object mismatches.

Viewed within this framework, the present findings suggest that manipulating task demands can shift the relative weighting of object detection and object identification. In particular, introducing a key feature of conventional visual search tasks to a change detection task likely produced a shift in process weighting from detection to identification. Specifically, knowledge of the target object prior to presentation of the scene, a task feature common in visual search, seems crucial in shifting the processing emphasis to one that benefits from the cohesive scene structure inherent to congruent context trials. Many of the

studies reporting congruency benefits in scene perception employ tasks that give the viewer target information prior to the presentation of the scene (Henderson et al., 1999; Neider & Zelinsky, 2006; Torralba et al., 2006; Vo & Henderson, 2011; Vo & Schneider, 2010). Knowledge of the upcoming target object, in conjunction with gist-based expectancies, might well lead to efficient scene exploration in search of a congruent target object.

In the current study, however, introducing target information prior to presentation of the scene did not produce congruency benefits when the task could also be interpreted as one that required change detection; that is, when scene presentations were brief and interrupted and the target appeared only in alternating scene presentations. Although the incongruency benefit in such tasks was reduced when participants were given target information at the beginning of each trial, performance remained more efficient for incongruent than for congruent trials. Brief and interrupted scene presentations may play a critical role in quite a few of the studies that have reported incongruency benefits (Hollingworth & Henderson, 2000; 2007; Hollingworth et al., 2001; LaPointe et al., 2013; LaPointe & Milliken, 2016). For example, in an experiment that removed the intervening white screens from between scene presentations in a change detection task, LaPointe et al. reported a congruency benefit rather than a incongruency benefit. This result converges with that of Experiment 3 of the current study, where presenting the scenes for extended and uninterrupted viewing in the context of a conventional visual search task eliminated the incongruency benefit and produced a trend toward a congruency benefit. These results

suggest that brief and interrupted scene presentations leave the viewer especially sensitive to violations of scene context.

At the same time, incongruency benefits have also been observed in studies that have used single scene presentations that were either extended (Loftus & Mackworth, 1978) or brief (Gordon, 2004) and that did not define the target as appearing and disappearing in alternating scenes. The findings from these studies suggest that brief and interrupted scene presentation on its own is not unique in promoting the type of processing that produces context incongruency benefits in scene perception. Rather, the current results suggest that the task requirement to search for a changing target under conditions of brief and interrupted scene presentation leads performance to depend heavily on shifts of attention to areas of the scene that require further interpretation (e.g., violations of expectancies). This task set is in contrast to situations in which performance depends heavily on what the scene context suggests ought to be present. Therefore, the key contrast across tasks producing congruency benefits and those producing incongruency benefits may be reliance on what the scene context suggests ought to be in the scene versus investigating what is actually in the scene. A possibility worth pursuing is that these two modes of processing map onto already well-established task set effects on attention capture. By this view, discrepancies between the target and scene context may hold the potential to capture attention, but whether that attention capture contributes to performance may ultimately depend on task factors that dictate top-down settings (Folk, Remington & Johnston, 1992; Bacon & Egeth, 1994).

In summary, to our knowledge, this is the first study to conduct a systematic task analysis of context congruency effects in visual scene perception; context congruency effects in visual search and change detection tasks were compared while controlling the stimulus set. The results help to reconcile contrasting results reported in the literature. Specifically, a number of studies have reported congruency benefits during scene perception (e.g., Davenport & Potter, 2004; Neider & Zelinsky, 2006; Palmer, 1975; Sun et al., 2011), whereas several others have reported incongruency benefits (Gordon, 2004; Hollingworth & Henderson, 2000; LaPointe et al., 2013; Loftus & Mackworth, 1978). Both the quality and complexity of the stimulus sets, as well as the task demands have varied widely across these studies. The results from the current study suggest that changes in the task alone can shift context congruency effects in opposing directions. This set of results is broadly consistent with the view that context congruency interferes with some processes and benefits others, and that the overall effect of context congruency on performance in scene perception tasks will depend on the specific weightings of these processes in any given task.

Footnote

Note 1. Perceptual salience was computed in two different ways, target salience and target change salience, using the same algorithm and process used in the first experiment (Zhang et al., 2008). There was no difference in terms of target salience for congruent (M = 137.17) and incongruent (M = 132.90) images, t (62) = .35, p = .73. There was also no difference in terms of target change salience for congruent (M = 15.12) and incongruent (M = 10.36) images, t (62) = .49, p = .63.

Chapter 5: General Discussion

Despite our perception of a veridical, complete, and seamless visual world, a large body of research has demonstrated that our representations are often inaccurate, incomplete, and created in a piecemeal fashion. Much of this research has focused on the perception and representation of complex scenes. Given the piecemeal nature with which we construct representations of our visual world, an important topic of research in complex visual scene processing concerns attentional prioritization of the components of a visual scene. What areas of a complex scene receive attentional priority, and what is the nature of this attention allocation?

Contradictory findings have obscured the research on complex scene perception. For example, many studies have found that the context of a scene can promote efficient perception of that scene by creating expectations of what objects might be encountered (Castelhano & Henderson, 2008; Palmer, 1975), by guiding attention to relevant areas (Neider & Zelinsky, 2006; Rensink, O'Regan, & Clark, 1997), and supporting the identification of local objects (Davenport & Potter, 2004; Sun, Simon-Dack, Gordon, & Teder, 2011). The results across these studies seem to suggest that a congruent context is important in guiding attention from one area of a scene to the next in support of a coherent representation. Yet, a growing list of studies has found that an incongruent context leads to efficient performance (e.g., Hollingworth & Henderson, 2000). A goal of the present research was to reconcile these conflicting results. To do so, three separate strategies were employed: (1) identify the processes that underlie the perception of complex scenes

(Chapter 2), (2) demonstrate that the relative weightings of these processes can be manipulated in such a way as to produce either congruency benefits or incongruency benefits(Chapters 2 & 4), and (3) investigate the nature of attention allocation under conditions in which behaviour supports incongruency benefitss (Chapter 3).

Multiple Processes Underlie Scene Perception

Across the numerous studies investigating the influence of context on scene perception, conflicting results have emerged both within (Henderson, Weeks, & Hollingworth, 1999; Loftus & Mackworth, 1978) and across tasks (e.g., Gordon, 2004; Vo & Henderson, 2011). There has been, however, one task that appears to produce incongruency benefits reliably - change detection (Bonitz & Gordon, 2008; Hollingworth & Henderson, 2000; 2007). In Chapter 2, a change detection task was used to replicate the robust finding of performance benefits on trials in which an incongruent context is presented. Further, by manipulating parameters of the change detection task, a benefit of context congruency was revealed. Specifically, removing the white screens typically interleaved between scene presentations produced faster response times for congruent than incongruent objects.

The results from Chapter 2 were taken as an indication that multiple processes underlie performance when viewing scenes for change. In particular, it was speculated that one such process, object detection, benefits from a disfluency between an object and the scene context. In contrast, another process, object identification, benefits from the structure inherent in a congruent context. The relative weightings of these processes, and subsequent

behaviour, will depend on the specific constraints of the task. In some circumstances, the task may be weighted heavily toward the process that produces congruency benefits, whereas in other circumstances the task is weighted more heavily toward the process that produces incongruency benefits.

Influence of Task Parameters on Context Effects

Although change detection tasks have routinely been used to produce incongruency benefits, visual search tasks have routinely been used to produce congruency benefits (although see Loftus & Mackworth, 1978). A common feature in many visual search tasks is to present a target object prior to the presentation of a complex scene. The goal of the participant is find the target object in the scene as quickly and accurately as possible. Under these constraints, participants often show performance benefits for objects embedded in a congruent context relative to objects embedded in an incongruent context. There are several key differences between visual search and change detection tasks. One such difference involves giving the viewer information about the target object prior to the presentation of the scene. It may be that having prior knowledge of the target object motivates viewers to use the context in search of the target. For example, seeing a kettle prior to the presentation of a kitchen scene, may lead attention towards areas of the kitchen scene in which a kettle would typically be found. In contrast, seeing a kettle prior to the presentation of a forest scene would be less helpful as the context contains little information that could be used to guide attention.

Chapter 4 describes a further demonstration of the multiple processes that underlie scene perception. Specifically, it was reasoned that knowledge of the target prior to the presentation of a complex scene should motivate viewers to use the structure inherent in a congruent context to efficiently search for and detect a target object. In contrast, knowledge of the target object prior to the presentation of an incongruent scene should afford no such benefit for object search and detection. In this case, participants were given an image or label describing the target object prior to the presentation of a congruent or incongruent scene using a change detection task. With these task parameters, the incongruency benefit typically measured in change detection tasks was significantly reduced. In a subsequent experiment, performance in a modified change detection task was directly compared to performance in a traditional visual search task while keeping the stimulus constant across tasks. In this case, it was demonstrated that task parameters can have profound effects on scene perception. Specifically, participants showed a trend toward a performance benefit for congruent objects when performing a visual search task, but a significant performance benefit for incongruent objects when performing a change detection task.

Influence of Stimulus Set Qualities on Context Effects

The results from Chapters 2 and 4 demonstrate that multiple processes underlie the perception of complex scenes. Across tasks, the relative weightings of these processes may very well produce congruency benefits in one situation and incongruency benefits in another. Indeed, this task weighting proposal may in part explain the conflicting results reported across studies using different tasks. However, it does not explain the conflicting

results that have been reported across studies using the same task. For example, Loftus and Mackworth (1978) asked participants to view a set of complex line drawings for 4 s each in preparation for a later scene recognition memory test. These researchers reported that incongruent target objects were likely to be fixated earlier into scene processing than congruent targets. Moreover, eye movements entering the target area were larger for incongruent than congruent targets. However, using a similar task, Henderson et al. (1999) reported different results. Here, participants showed no difference in the early eye fixations to congruent relative to incongruent targets.

In both the study by Loftus and Mackworth (1978) and that of Henderson et al. (1999) complex line drawings were presented for extended viewing in preparation for a later recognition memory test. In one study, incongruency benefits were measured (Loftus & Mackworth, 1978), whereas in the other study incongruency benefits were not measured (Henderson et al., 1999). In both studies, a stimulus set composed of complex black and white line drawings were used. Despite the seemingly similar stimulus sets, the images used by Loftus and Mackworth have been criticized for lacking complexity compared with stimulus sets used by others (Henderson et al., 1999; Vo & Henderson, 2011), perhaps leading to the incongruency benefits reported in that study. As a consequence of the image simplicity, the semantically incongruent target objects in the stimulus set used by Loftus and Mackworth may have been more perceptually salient relative to the other objects in the scenes. As such, it may have been perceptual factors, rather than semantic factors, driving performance in that study.

Despite attempts to control the perceptual salience of the target objects across the context conditions in Chapters 3 and 4[1], the results from the second and third experiments in Chapter 4 are an example of how stimulus set characteristics can influence behaviour when perceiving complex scenes. A subset of the images used in the second experiment of Chapter 4 contained distractor objects that resembled the target object both in terms of semantic and perceptual characteristics. For example, an image containing a target chair might also contain other chairs as distractor objects. When comparing performance in the second experiment of Chapter 4 to performance in the third experiment, where the images containing multiple tokens of the target object were removed, the multiple tokens appear to slow response times. This issue seemed especially prevalent in the congruent context condition. The difference in performance across these two experiments further demonstrates that stimulus characteristics can have measurable effects on task performance.

The Nature of Attention Allocation

The results reported in Chapters 2 and 4 demonstrate that multiple processes underlie scene perception and the relative weighting of these processes can produce congruency benefits in some cases and incongruency benefits in others. Moreover, the results from these experiments demonstrate that task parameters and stimulus qualities can influence the relative weighting of these processes and impact behaviour. However, the question remains, when perception is biased towards a process that benefits from either a congruent or incongruent context, what is the nature of attention allocation? This question

has received some speculation in the literature, specifically in regards to the incongruency benefits routinely measured in change detection tasks. Hollingworth and Henderson (2000) argued that change detection tasks may produce superior performance for incongruent objects than congruent objects in one of two ways. It may be that attention is drawn to the incongruent objects as the identity of the object conflicts with the semantic information contained in the context of the scene. The authors termed this the attention attraction hypothesis. Conversely, attention may be dispersed about the scene in a random way. However, once attention has landed on an incongruent object there may be a tendency for it to linger, perhaps in an effort to identify the oddball object. The authors termed this the attention disengagement hypothesis.

The hypotheses submitted by Hollingworth and Henderson (2000) are compelling in that the attention attraction hypothesis would seem to violate a basic tenet of early selection models of attention (Broadbent, 1958). Namely, according to the attention attraction hypothesis, the identity of the incongruent object draws attention, the implication being that semantic information associated with the object is recruited pre-attentively. This assumption is a direct violation of early selection models in that according to these models attention is required to extract meaning. Conversely, the attention disengagement hypothesis does not violate early selection models of attention. According to this hypothesis, attention is dispersed randomly about the scene in a serial fashion. It is not until attention has landed on an object that meaning is extracted. In fact, according to this hypothesis, the reason attention lingers on an incongruent object is because meaning is yet

to be extracted. It is also important to note that the hypotheses offered by Hollingworth and Henderson are not mutually exclusive. It is conceivable that attention is quickly drawn to incongruent objects and lingers on these objects for further processing.

The purpose of study reported in Chapter 3 was to investigate the nature of attention allocation under conditions in which incongruency benefits can be measured. To do so, participants completed a change detection task while their eye movements were measured. The strategy here was to investigate the pattern of eye movements early during scene processing. If eye movements tend to move in the direction of incongruent objects preferentially and early into scene processing, this would be taken as evidence that semantically incongruent objects attract attention. Conversely, if eye movements show no bias towards incongruent objects early during scene processing, but have a tendency to linger on these objects, this would be taken as evidence that semantically incongruent objects do not attract attention, but rather fail to disengage from these objects. The results from this experiment demonstrate that attention is captured by semantically incongruent object early into scene processing. Moreover, the results offer no evidence that attention lingers on incongruent objects any more than it does for congruent objects. From these results, it was argued that attention is attracted to semantically incongruent objects early when viewing scenes for change.

The findings of the Chapter 3 appear to be a clear violation of early selection models of attention. Specifically, the results indicate that objects that are semantically incongruent with the scene context capture attention early into scene processing. These

results might be taken as an indication that incongruency at the semantic level of processing attracts attention. However, the link between semantic processing incongruency and attention capture in this study is debatable. Although the objects were matched in terms of perceptual salience across the context condition, note that semantically incongruent objects are unlikely to have been perceptually processed in the past with the other objects in the scene. That is, semantically incongruent target objects are poorly predicted at all levels of processing by the context in which they occur, and poor prediction at the perceptual level could have negative consequence for the unfolding of perceptual processing of semantically incongruent targets. Disfluency in processing at a perceptual level, therefore, could provide a basis for attention capture by semantically incongruent objects.

To clarify this point, consider an experiment by Gordon (2006). Participants were given a brief presentation of a line drawing of a complex scene (100 ms), followed by a pattern mask (10 ms), a blank screen (100 ms), and finally a letter string. Participants were tasked with making a lexical decision for the letter string, indicating whether or not it formed a legal English word. Unbeknownst to the participants, the legal letter strings described an object that was congruent and either present or not present in the preceding scene, or an object that was incongruent and either present or not present in the preceding scene. Gordon reasoned that if the semantic characteristics of the semantically incongruent objects had been processed during the brief exposure of the scene, participants should be more likely to respond accurately on the letter strings that described these objects, relative

to the letter strings that described incongruent objects that were not present in the same image (i.e., positive priming). However, Gordon did not find positive priming for incongruent-present letter strings. Instead he found negative priming for the letter strings describing congruent-present letter strings. From these results, Gordon argued that attention was indeed attracted to the incongruent targets and drawn away from the congruent targets, creating negative priming for the congruent-present letter strings. However, the nature of the attention attraction was not due to the semantic characteristics of the incongruent object, therefore no positive priming was measured for the incongruent-present letter strings. Instead, the perceptual characteristics of the incongruent target drew attention.

The results reported by Gordon (2006) cast some doubt on the view that the attraction of attention to incongruent objects reported in Chapter 3 were due to semantic characteristics of the objects. Rather than semantic characteristics, argues Gordon, it is the unfolding of perceptual processing of the object that captures attention. As the viewer quickly recruits the contextual information of the scene, perceptual expectations are formed. When the scene contains perceptual features that are not normally associated with that scene context (e.g., semantically incongruent objects), attention is attracted to that region of the scene. In this way, perceptual disfluency may capture attention, without any need to pre-attentively identify an object. This perspective does not necessarily violate tenets of early selection models.

Summary

The goal of the present research was to investigate the nature of processing during complex scene perception. The particular strategy used here was to focus on the influence of scene context on object processing. To this point, the literature has been riddled with contradictory findings, with some studies showing performance benefits for context congruency and others showing performance benefits for context incongruency. The current research was designed to reconcile these conflicting results using a three-pronged approach: (1) identify the processes that support scene perception; (2) manipulate the relative weightings of these processes in such a way as to produce either congruency benefits or incongruency benefits; and (3) investigate the nature of attention allocation while processing is biased towards incongruency benefits.

From the results, two processes have been identified that support the perception of complex scenes. One process, object identification, appears to benefit from a congruent context; whereas, another process, object detection, appears to benefit from an incongruent context. Moreover, the current results appear to support the notion that the relative weightings of the processes produce congruency benefits in some cases, and incongruency benefits in others. The findings may help explain the conflicting results that have been reported across tasks in the literature. The current results also point to the importance of stimulus qualities and their influence on congruency effects. Certain stimulus qualities may push processing in support of congruency benefits, whereas others may push processing in support of incongruency benefits. This framework may help reconcile the conflicting

results that have been reported across studies using similar tasks. Finally, the current research suggests that in situations in which processing is biased in favour of incongruency benefits, attention is attracted to semantically incongruent objects. Together, the results of the current research help reconcile several long standing issues of debate in the scene perception literature.

The current research, however, also leaves several questions unanswered. For example, when processing is biased in favour of congruency benefits, what is the nature of attention allocation? It may be that attention attraction to semantically congruent objects is fundamentally similar to attention attraction to semantically incongruent objects. On the other hand, it may be that attention to congruent objects reflects a serial deployment of attention that is patterned; that is, serial shifts of attention that are constrained and guided by the scene context from one fixation to the next. In contrast, attention to incongruent objects may reflect discrete shifts of attention to areas of a scene where further meaning is required. Another outstanding issue concerns the nature of attention capture when processing is biased in favour of incongruency benefits. The current research suggests that attention is attracted to semantically incongruent objects when viewing scenes for change, but there are in fact several hypotheses that offer competing explanations of precisely how that attention capture occurs. For instance, it may be that the conflict in identity between a semantically incongruent object and the scene context attracts attention. On the other hand, it may be that conflict in expectations set out by the scene context and disfluent perceptual processing of a semantically incongruent object attracts attention. Finally, the current

research does not address how the processes that support scene perception might operate in other task domains. Using the structure inherent in the real-world (e.g., congruent contexts) would seem fundamental to supporting efficient behaviour in a wide array of task domains. Yet, the ability to attend to novelty (e.g., incongruent objects) and learn about structure that is not well represented seems like it would be equally important in a wide array of task domains. If these two broad principles are at work in complex scene perception but also in other task task domains, an important goal for future research is to examine the portability of this processing distinction across various task domains.

Footnote

Note 1. The salience analysis I chose uses a combination of difference of Gaussians (DoG) filters on the intensity and colour channels to compute degrees of rarity for each pixel in the image (Zhang, Tong, Marks, Shan, & Cottrell, 2008). The use of this local informativeness measure is different than more traditional salience models that calculate global salience (Bruce & Tsotsos, 2006; Gao & Vasconcelos, 2007; Itti, Koch, & Nibur, 1998; Torralba, Oliva, Castelhano, & Henderson, 2006). The advantage of computing local salience, in contrast with global salience, is a more efficient and biologically relevant model. Moreover, the salience model used in this thesis has been shown to perform as well or better than a number of other salience models at predicting eye movements from an independent group of human observers (Bruce & Tsotsos, 2006, Gao & Vasconcelos, 2007; Itti et al, 1998, Zhang et al., 2008).

References

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

Bacon, W. T., & Egeth, H. E. (1994). Overriding stimulus-driven attentional capture. *Perception and Psychophysics, 55*, 485-496.

Becker, M. W., Pashler, H., & Lubin, J. (2007). Object-intrinsic oddities draw early saccades. *Journal of Experimental Psychology: Human Perception and Performance, 35*, 20-30.

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

Biederman, I. (1981). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94*, 115-147.

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

Bonitz, V. S., & Gordon, R. D. (2008). Attention to smoking-related and incongruous objects during scene viewing. *Acta Psychologica, 129*, 255-263.

Boyce, S. J., Pollatsek, A., Rayner, K. (1989). Effect of background information on object identification. *Journal of Experimental Psychology: Human Perception and Performance, 15*, 556-566.

Bracco, F., & Chiorri, C. (2009). People have the power: Priority of socially relevant stimuli in a change detection task. *Cognitive Processes, 10*, 41-49.

Brand, J. (1971). Classification without identification in visual search. *Quarterly Journal of Experimental Psychology, 23*, 178-186.

Broadbent, D. E. (1958). *Perception and communication.* New York: Oxford University Press.

Brockmole, J. R., & Henderson, J. M. (2008). Prioritizing new objects for eye fixation in real-world scenes: Effects of object-scene consistency. *Visual Cognition, 16*, 375-390.

Bruce, N., & Tsotsos, J. (2006). Saliency based on information maximization. In Y. Weiss, B. Scholkopf, & J. Platt (Eds.), *Advances in neural information processing systems 18* (pp. 155-162). Cambridge, MA: MIT Press.

Buswell, G. T. (1935). *How people look at pictures.* Chicago: University of Chicago Press.

Castelhano, M. S., & Henderson, J. M. (2008). The influence of color on the perception of scene gist. *Journal of Experimental Psychology: Human Perception and Performance, 34*, 660-675.

Chun, M. M. (2003). Scene perception and memory. In D. E. Irwin & B. H. Ross (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 42, pp. 79-108). San Diego, CA: Academic Press.

Chun, M. M., & Jiang, Y. (1998). contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology, 36*, 28-71.

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology, 1*, 42-45.

Curran, T., Gibson, L., Horne, J. H., Young, B., & Bozell, A. P. (2009). Expert image analysts show enhanced visual processing in change detection. *Psychonomic Bulletin and Review, 16*, 390-397.

Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science, 15*, 559-564.

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

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

Duncan, J. (1983). Category effects in visual search: A failure to replicate the "oh-zero" phenomenon. *Perception & Psychophysics, 34*, 221-232.

Fabre-Thorpe, M., Delorme, A., Marlot, C., & Thorpe, S. J. (2001). A limit to the speed of processing in ultra-rapid visual categorization of novel natural scenes. *Journal of Cognitive Neuroscience, 13*, 171-180.

Fei-Fei, L., Iyer, A., Koch, C., & Perona, P. (2007). What do we perceive in a glance of a real-world scene? *Journal of Vision, 7*, 1-29.

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*, 1030-1044.

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

Gao, D., & Vasconcelos, N. (2007). Bottom-up saliency is a discriminant process. In *IEEE International Conference on Computer Vision (ICCV'07)*. Rio de Janeiro, Brazil.

Gareze, L., & Findlay, J. M. (2007). Absence of scene context effects in object detection and eye gaze capture. In R. van Gompel, M. Fischer, W. Murray, & R. W. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 537-562). Amsterdam: Elsevier.

Gibson, J. J. (1979). *The ecological approach to visual perception.* Boston: Houghton Mifflin Company.

Gordon, R. D. (2004). Attentional allocation during the perception of scenes. *Journal of Experimental Psychology: Human Perception and Performance, 30*, 760-777.

Gordon, R. D. (2006). Selective attention during scene perception: Evidence from negative priming. *Memory and Cognition, 34*, 1484-1494.

Green, D., & Hummel, J. E. (2006). Familiar interacting object pairs are perceptually grouped. *Journal of Experimental Psychology: Human Perception and Performance, 32*, 1107-1119.

Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences, 9*, 188-193.

Henderson, J. M., Brockmole, J. R., Castelhano, M. S., & Mack, M. (2007). visual saliency does not account for eye movements during visual search in real-world scenes. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds). *Eye movements: A window on mind and brain* (pp. 538-562). Amsterdam: Elsevier.

Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology, 50*, 243-271.

Henderson, J. M., Weeks, P. A., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance, 25*, 210-228.

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

Hollingworth, A., & Henderson, J. M. (2007). *Testing conceptual and visual short-term memory explanations for the inconsistent object change detection advantage in real-world scenes.*Unpublished technical report. Eyelab: Michigan State University.

Hollingworth, A., Schrock, G., & Henderson, J. M. (2001). Change detection in the flicker paradigm: The role of fixation position within the scene. *Memory and Cognition, 29*, 296-304.

Hollingworth, A., Williams, C. C., & Henderson, J. M. (2001). To see and remember: Visually specific information is retained in memory from previously attended objects in natural scenes. *Psychonomic Bulletin and Review, 8*, 761-768.

Intraub, H. (1980). Presentation rate and the representation of briefly glimpsed pictures in memory. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 1-12.

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

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40*, 1489-1506.

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20*, 1254-1259.

Jones, B. T., Jones, B. C., Smith, H., & Copley, N. (2003). A flicker paradigm for inducing change blindness reveals alcohol and cannabis information processing biases in social users. *Addiction, 98*, 235-244.

Jonides, J., & Gleitman, H. (1972). A conceptual category effect in visual search: O as letter or as digit. *Perception & Psychophysics, 12*, 457-460.

Joubert, O. R., Rousselet, G. A., Fize, D., & Fabre-Thorpe, M. (2007). Processing scene context: Fast categorization and object interference. *Vision Research, 47*, 3286-3297.

Kelley, T., Chun, M., & Chua, K. (2003). Effects of scene inversion on change detection of targets matched for visual salience. *Journal of Vision, 3*, 1-5.

Lachter, J., Forster, K. I., & Ruthruff, E. (2004). Forty-five years after Broadbent (1958): Still no identification without attention. *Psychological Review, 111*, 880-913.

Lachter, J., Ruthruff, E., Lien, M.-C., & McCann, R. S. (2008). Is attention needed for word identification? Evidence from the Stroop paradigm. *Psychonomic Bulletin & Review, 15*, 950-955.

LaPointe, M. (2011). *Testing the animate monitoring hypothesis* (Unpublished master's thesis). University of Lethbridge, Lethbridge, Alberta, Canada.

LaPointe, M. R. P., Lupiáñez, J., & Milliken, B. (2013). Context congruency effects in change detection: Opposing effects on detection and identification. *Visual Cognition, 21*, 99-122.

LaPointe, M. R. P., & Milliken, B. (2016). Semantically incongruent objects attract eye-gaze when viewing scenes for change. *Visual Cognition.* Online first publication. DOI: 10.1080/13506285.2016.1185070

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

Lupiáñez, J., Ruz, M., Funes, M. J., & Milliken, B. (2007). The manifestation of attentional capture: Facilitation or IOR depending on task demands. *Psychological Research, 71*, 77-91.

Mack, A., & Rock, I. (1998). *Inattentional blindness* (Vol. 33). Cambridge, MA: MIT press.

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

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

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

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

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorial in Quantitative Methods for Psychology, 4*, 61-64.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology, 9*, 353-383.

Neider, M. B., & Zelinsky, G. J. (2006). Scene context guides eye movements during visual search. *Vision Research, 46*, 614-621.

Neisser, U., & Becklen, R. (1975). Selective looking: Attending to visually specified events. *Cognitive Psychology, 7*, 480-494.

Oliva, A., & Schyns, P. (1997). Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognitive Psychology, 34*, 72-107.

Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *Trends in Cognitive Science, 11*, 520-527.

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

Pezdek, K., Whetstone, T., Reynolds, K., Askari, N., & Dougherty, T. (1989). Memory for real-world scenes: The role of consistency with schema expectation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 587-595.

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

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory, 2*, 509-522.

Potter, M. C., & Levy, E. I. (1969). Recognition memory for a rapid sequence of pictures. *Journal of Experimental Psychology, 81*, 10.

Rayner, K., Castelhano, M. S., & Yang, J. (2009). Viewing task influences eye movements during active scene perception. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 35*, 254-259.

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*, 368-373.

Ruz, M., & Lupiáñez, J. (2002). A review of attentional capture: On its automaticity and sensitivity to endogenous control. *Psicologica, 23*, 283-309.

Sampanes, A. C., Tseng, P., & Bridgeman, B. (2008). The role of gist in scene recognition. *Vision Research, 48*, 2275-2283.

Schyns, P., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time- and spatial-scale dependent scene recognition. *Psychological Science, 5*, 195-200.

Smilek, D., Dixon, M. J., & Merikle, P. M. (2006). Revisiting the category effect: The influence of meaning and search strategy on the efficiency of visual search. *Brain Research, 1080*, 73-90.

Simons, D. J., & Chabris, C. F. (1999). Goriallas in our midst: Sustained inattentional blindness for dynamic events. *Perception, 28*, 1059-1074.

Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences, 1*, 261-267.

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

Stirk, J. A., & Underwood, G. (2007). Low-level visual saliency does not predict change detection in natural scenes. *Journal of Vision, 7*, 1-10.

Sun, H.-M., Simon-Dack, S. L., Gordon, R. D., & Teder, W. A. (2011). Contextual influences of rapid object categorization in natural scenes. *Brain Research, 1398*, 40-54.

Theeuwes, J. (1991). Exogenous and endogenous control of attention: The effect of visual onsets and offsets. *Perception & Psychophysics, 49*, 83-90.

Theeuwes, J. (1992). Perceptual selectivity for color and form. *Perception & Psychophysics, 51*, 599-606.

Theeuwes, J. (1994a). Endogenous and exogenous control of visual selection. *Perception, 23*, 429-440.

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

Torralba, A., Oliva, A., Castelhana, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review, 113*, 766-786.

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

Treisman, A., Sykes, M., & Gelade, G. (1977). Selective attention and stimulus integration. *Attention and Performance, VI*, 333-361.

Turatto, M., & Galfana, G. (2000). Color, form and luminance capture attention in visual search. *Vision Research, 40*, 1639-1643.

Underwood, G., & Foulsham, T. (2006). Visual saliency and semantic incongruency influence eye movements when inspecting pictures. *Quarterly Journal of Experimental Psychology, 59*, 1931-1949.

Underwood, G., Humphreys, L., & Cross, E. (2007). congruency, saliency, and gist in the inspection of objects in natural scenes. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 564-579). Amsterdam: Elsevier.

Underwood, G., Templeman, E., Lamming, L., & Foulsham, T. (2008). Is attention necessary for object identification? Evidence from eye movements during the inspection of real-world scenes. *Consciousness and Cognition, 17*, 159-170.

VanRullen, R., & Thorpe, S. J. (2001). The time course of visual processing: From early perception to decision-making. *Journal of Cognitive Neuroscience, 13*, 454-461.

Vo, M. L.-H., & Henderson, J. M. (2009). Does gravity matter? Effects of semantic and syntactic inconsistencies on the allocation of attention during scene perception. *Journal of Vision, 9*, 1-15.

Vo, M. L.-H., & Henderson, J. M. (2011). Object-scene inconsistencies do not capture gaze: Evidence from the flash-preview moving-window paradigm. *Attention, Perception, & Psychophysics, 73*, 1742-1753.

Vo, M. L.-H., & Schneider, W. X. (2010). A glimpse is not a glimpse: Differential processing of flashed scene previews leads to differential target search benefits. *Visual Cognition, 18*, 171-200.

Weaver, M. D., & Lauwereyns, J. (2011). Attentional capture and hold: The oculomotor correlates of the change detection advantage for faces. *Psychological Research, 75,* 10-23.

Werner, S., & Thies, B. (2000). Is "change blindness" attenuated by domain-specific expertise? An expert-novices comparison of change detection in football images. *Visual Cognition, 7*, 163-173.

Yantis, S., & Jonides, J. (1990). Abrupt visual onsets and selective attention: Evidence from visual search. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 601-621.

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

Zhang, L., Tong, M. H., Marks, T. K., Shan, H., & Cottrell, G. W. (2008). SUN: A Bayesian framework for saliency using natural statistics. *Journal of Vision, 8*, 1-20.

Zoest, W. V., Donk, M., & Theeuwes, J. (2004). The role of stimulus-driven and goal-driven control in saccadic visual selection. *Journal of Experimental Psychology: Human Perception and Performance, 30*, 746-759.