The effects of the allocation of attention on rapid scene categorization

John O. Brand

A Thesis in the

Department of Psychology

Presented in Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy at

Concordia University

Montréal, Québec, Canada

November 2014

CONCORDIA UNIVERSITY
SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared

By: John O. Brand

Entitled: The effects of the allocation of attention on rapid scene categorization

and submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy (Psychology)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

____________________ Chair
Dr. G. Brown

____________________ External Examiner
Dr. M. Castelhano

____________________ External to Program
Dr. B. Grohmann

____________________ Examiner
Dr. R. Gurnsey

____________________ Examiner
Dr. R. de Almeida

____________________ Thesis Supervisor
Dr. A. Johnson

Approved by: ____________________
Dr. A. Arvanitogiannis, Graduate Program Director

November 7, 2014 ____________________
Dr. A. Roy, Dean, Faculty of Arts & Science

ABSTRACT

The effects of the allocation of attention on rapid scene categorization

John O. Brand, Ph.D.
Concordia University, 2014

It is well documented that observers are able to accurately extract the semantic information from natural scenes in 120 msec (Thorpe, Fize, & Merlot, 1996). This rapid categorization ability is often cited as evidence that the information that is required to categorize a scene originates from low-level visual information. Information related to an image's spatial scales (Oliva & Schyns, 1997; Schyns & Oliva, 1994), phase (Joubert, Rousselet, Fabre-Thorpe, & Fize, 2009; Loschky et al., 2007, 2010; Loschky & Larson, 2008), overall summary statistics (Evans & Treisman, 2005), and colour (Castelhano & Henderson, 2008; 2005; Loschky & Simons, 2004; Oliva & Schyns, 2000) have all been shown to provide information that can be used to categorize a briefly presented image. The experiments reported in this dissertation were designed to address the overarching question of how the visual system selects diagnostic scene information? It addressed this question by examining the hypothesis that visual attention facilitates the selection of information that underpins rapid scene categorization. In order to investigate this hypothesis, the present work was divided into two main manuscripts. Manuscript 1 is presented in Chapter 2 and includes four experiments that were designed to investigate if attending to global and local levels of a scene facilitate categorization based on a scene's coarse and fine information, respectively. This hypothesis was explored by asking observers to classify hybrid images. A hybrid image combines the coarse information (conveyed by an image's low spatial frequencies) of one image (e.g., a city) and the fine information (conveyed by an image's high spatial frequencies) of a second image (e.g., a highway). Experiments 1 and 2 showed that

although observers could classify hybrid images based on both fine and coarse information (i.e., as either a city or a highway scene; Experiment 1), observers preferred to base categorization on coarse content (Experiment 2). Experiment 3 demonstrated that categorization based on coarse content was facilitated when observers were prompted to attend globally to scenes compared to when they were prompted to attend locally. Experiment 4 demonstrated that this global facilitation effect was due, in part, to the facilitation of a hybrid's low spatial frequencies.

Manuscript 2 is presented in Chapter 3 and contains four experiments that investigated the hypothesis that distributed attention facilitates the extraction of a scene's overall summary statistics, which in turn, facilitates the ability to rapidly categorize scenes (Evans & Treisman, 2005). This hypothesis was investigated by examining whether manipulations of attention affected scene categorization in the same fashion as the extraction of overall summary statistics. Experiment 1 replicated the result that extraction of a scene's summary statistics is more compatible with distributed attention than focused attention (Chong & Treisman, 2005). Experiments 2 and 4 extended this finding by demonstrating that superordinate level categorization of both animals (e.g., detect the presence [or absence] of an animal, Experiment 2), and natural scenes (e.g., was the scene natural? Experiment 4), were more compatible with distributed than focused attention. However, Experiment 3 showed that there was no difference between the effects of distributed and focused attention on basic level categorization (e.g., was this a beach scene?).

Together, the findings of this thesis demonstrate that visual attention is important in the rapid categorization of a natural scene, by facilitating the selection of scene information that is necessary to classify a scene category.

Acknowledgements

First and foremost I would like to thank Dr. Aaron Johnson for his guidance and supervision over the past four years. I will be forever grateful for his willingness to supervise my research, and for his patience, guidance, enthusiasm, and expertise. This thesis would not have been possible without his leadership. I am also grateful to the late Dr. Michael von Grünau, whose passion and enthusiasm to share his expertise provided the foundation on which this thesis is based. I wish to thank my family for providing unconditional support and encouragement: My mother, Jill Pardy; my sister, Jennifer Lowe; and my grandmother, Pearl Paulson. I wish to honour the memory of my father, Owen Brand and my Grandfather, Leslie Pardy. I wish to also thank Sivan Rotenberg for her continuous support and incredible patience throughout this entire process.

The completion of this thesis would not have been possible without the financial assistance of my supervisor, Dr. Aaron Johnson, who through his NSERC grant was able to fund this project in the form of a research stipend. I would like to thank the Government of Quebec for funding this project in the form of a Fonds de recherche Nature et technologies du Québec doctoral scholarship. I offer my sincerest gratitude to my committee members, Dr. Rick Gurnsey and Dr. Roberto de Almeida for their willingness to share their knowledge.

CONTRIBUTIONS OF AUTHOR

This dissertation consists of a general introduction, two manuscripts (each consisting of four studies for a total of eight experiments) and a concluding chapter. I wrote the general introduction and concluding chapter with feedback from my supervisor, Dr. Aaron Johnson. The contributions of Dr. Aaron Johnson and myself to the co-authored manuscripts reported in this dissertation are discussed below.

Chapter 2: Attention to local and global levels of hierarchical Navon figures affects rapid scene categorization.

In collaboration with Dr. Aaron Johnson, I designed the four experiments and was responsible for programming all of the experiments. All participants were recruited using the psychology participant poll at Concordia University and I along with undergraduate students, Yvette Esses, Jessica Wilson, and Diana Mihalache, were responsible for collecting the data. I conducted all statistical analyses and drafted the manuscript with feedback from Dr. Johnson. Furthermore, I revised the manuscript based on the comments from two anonymous reviewers after consulting with Dr. Johnson. Dr. Johnson provided the Matlab code that was used to create the scene stimuli used in the experiments reported in Chapter 2, and the images produced in Appendix 2.1 and 2.2.

Chapter 3: The effects of distributed and focused attention on rapid scene categorization

As in Manuscript 1, I designed all four experiments reported in Manuscript 2, in collaboration with Dr. Aaron Johnson. I programmed Experiments 1, 2 and 4 and Dr. Johnson programmed Experiment 3. Participants were recruited using the psychology participant poll at Concordia University and I along with undergraduate students, Yvette Esses and Jessica Wilson, were responsible for data collection. I analyzed all data and drafted the manuscript with feedback from Dr. Johnson.

TABLE OF CONTENTS

LIST OF FIGURES

Figures in Chapter 1

Figures in Chapter 2

Figures in Chapter 3

Figures in Appendix A

LIST OF TABLES

Tables in Chapter 2

Tables in Chapter 3

Chapter 1: General Introduction

At any given second, the human retina is capable of transmitting approximately 10 million bits of information to the visual areas of the brain (i.e., roughly the equivalent of the transmission speed of a modern Ethernet connection; Koch, McLean, Segev, Freed, Berry II, Balasubramanian, & Sterling, 2006). However, rather than process this vast amount of information, the visual system evolved to select only a relevant subset for further processing. Classical theories of visual perception posit that only attended information will reach conscious awareness whereas unattended information, although still processed, will not be consciously perceived (Treisman & Gelade, 1980). Consequently, it is suggested that attention mediates visual perception by breaking it down into two distinct stages: a pre-attentive stage and an attentive stage (e.g., Feature Integration Theory of Attention, Treisman & Gelade, 1980; Guided Search Theory of Visual Search, Wolfe, Cave, & Franzel, 1989). The pre-attentive stage is rapid (less than 200 msec) and processes visual information in parallel across the entire visual field. During this stage, salient separable features (e.g., colour, size, and orientation) are automatically encoded onto separate feature maps that determine their location in visual space. In contrast to the pre-attentive stage, the attentive stage is slower – typically over 200 msec - and is responsible for binding together the features encoded in the pre-attentive stage. For example, each attended location in feature space results in the binding of separable features that appear at that location, resulting in the perception of a consciously experienced whole (e.g., attention is required in order to bind the features red and vertical into the perception of a red vertical item; see, e.g., Block, 2005; Koch & Tsuchiya, 2007; Lamme, 2003; O'Regan & Noë, 2001; Posner, 1994).

Although this conceptualization of visual attention is successful in describing experimental results related to basic geometric shapes (e.g., visual search; Wolfe et al., 1989), it is less successful in describing research related to complex stimuli (Braun, 2003). Observers are able to extract both statistical and semantic information from sets of similar objects and natural scenes, respectively, without the need to encode individual elements, suggesting that attention is not needed to bind features together. For example, observers can extract the average size of a set of circles (Ariely 2001, Chong & Treisman 2003; 2005), average direction of a set of randomly moving blots (Watamaniuk, Sekuler, & Williams, 1989), and average emotion or identity in a set of similar faces (de Fockert & Wolfenstein, 2009; Haberman, & Whitney, 2007, 2009; Haberman, Harp, & Whitney, 2009), without having a good representation of the individual items that comprise the set. Additionally, observers are able to extract semantic information from scenes presented within the time period thought to coincide with pre-attentive processing. Within 120 msec, observers are able categorize a natural scene according to either its basic (e.g., beach and desert) or superordinate (e.g., indoor and outdoor) level (Greene & Oliva, 2009; Joubert, Rousselet, Fize & Fabre-Thorpe, 2007), suggesting that initial scene comprehension occurs prior to this time point.

Findings related to the rapid extraction of statistical and semantic information are influential in forming contemporary theories about scene perception. Specifically, researchers have focused on explaining how some higher-level scene related behaviours could occur in the absence of attention (i.e., in the pre-attentive stage). In their influential work, Evans and Treisman (2005; see also Chong & Treisman, 2005; Treisman, 2006) suggested that the automatic and pre-attentive extraction of semantic and statistical information is facilitated by different attentional distributions. The theory is that there are different types of attention, and that each type is associated with the extraction of a different type of information. According to Evans and Treisman, the perception of our environment is organized hierarchically. For any given complex scene (e.g., a beach) attention can be directed globally, locally, or distributed over a set of similar items. When attention is focused locally on a particular object (e.g., a palm tree), attention acts to bind its separable features together, enabling object identification (see, e.g., Feature Integration Theory of Attention, Treisman & Gelade, 1980). In contrast, when attention is distributed over a set of similar items (e.g., a group of beach pebbles), the visual system automatically computes summary statistics related to the set (e.g., the average pebble size) without having a good representation of any individual set member. Finally, when attention is set globally to the scene as a whole, the visual system extracts multiple summary representations that act as a set of disjunctive features that can be used to extract a scene's meaning (e.g., a beach on a sunny day) without the need for focused attention (see Evans & Chong, 2011 for a review).

Similar to Evans and Treisman (2005), Oliva and Schyns (1997) proposed that different scene information is associated with different attentional distributions. However, whereas Evans and Treisman suggested that distributed attention facilitates the extraction of a scene's summary statistics, Oliva and Schyns argued that attending locally and globally facilitates the selection of a scene's fine and coarse scale information. Their hypothesis is that natural images are encoded by the visual system via different spatial frequency channels, and that each spatial frequency is associated with a different level of information. Consider, for example, the images presented in Figure 1.1. The leftmost and rightmost images have been filtered so that only their low, and high spatial frequencies remain. The image located in the centre is a full broadband image and contains all its spatial frequencies. As is evident in the figure, low spatial frequencies convey information related to an image's global properties (e.g., overall shape, such as general orientation), whereas high spatial frequencies convey information related to an image's local properties (e.g., configural and fine details). According to Oliva and Schyns, attention can be directed to either the local, or the global level; the attended level determines the information to be used as the basis for categorization.

Figure 1.1. An example of a low-pass filtered image, a broadband image, and a high-pass filtered image. The leftmost and rightmost images have been filtered, such that only their low, and high-spatial frequencies remain. The image in the centre has not been filtered and contains all spatial frequencies.

The overarching goal of the present dissertation is to address how attending locally and globally affects the selection of information used for rapid scene categorization. Whereas both Evans and Treisman (2005) and Oliva and Schyns (1997) argued that the rapid extraction of semantic meaning is dependent on the deployment of attention, they differ with respect to both the type of attention that is required, and the associated information that is used as the basis for categorization. Consequently, the present work is broken down into two main chapters. Chapter 2 addresses of how attention to local and global levels facilitates rapid scene categorization based on fine and coarse information, respectively (Oliva & Schyns, 1997), using filtered images. Chapter 3 addresses how adopting a distributed attention distribution facilitates the rapid extraction of summary statistics that can be used as the basis for rapid scene categorization (Evans & Treisman, 2005). In pursuit of the answers to these questions, both Chapters 2 and 3 present the results of four experiments, with each experiment being related to the overarching question of each chapter. Nevertheless, before explaining these experiments it is beneficial to first describe the research on which they are based**.** Thus, the questions that motivate each respective chapter are described in section 1.5 of the present introduction. The primary goal of the following sections is to describe the most relevant research as it pertains to the current investigations. Particular emphasis is placed on research examining the effects of attention on scene categorization, and how the results contribute to the development of each respective theory. Furthermore, a secondary goal of the introduction is to provide the necessary background in order to explain how the two theories potentially integrate. This integration is discussed in section 1.4.3.

1.1 Rapid scene categorization

In their seminal work, Potter and Levy (1969) addressed what contextual scene information is available at very brief presentation durations. They asked their observers to complete two phases: a scene memorization phase and a scene test phase. In the memorization phase, observers were asked to memorize a series of target images presented at 123 msec/image. Immediately following the memorization phase, observers subsequently completed the test phase in which they were asked to the identify the presence of target images either presented alone, or embedded in a rapid serial visual presentation stream (RSVP stream; 123, or 250 msec presentation duration for each image). When images were presented in isolation, recognition accuracy was high; however, when targets were embedded in an RSVP stream, target recognition regressed to chance performance, at both presentation durations. In attempt to better understand these findings, Potter (1976) conducted a follow-up experiment in which she replicated her original design with the following exception: rather than including a memorization phase, Potter cued the target category using either a visual, or verbal prime prior to the start of the RSVP stream (e.g., for the target category "beach', the target cue was either a picture of a beach scene, or the word beach). In contrast to her original report, target identification (collapsed over cue conditions) was above 60% and 80%, respectively, for presentation durations of 123 and 250 msec. Taken together, these two seminal papers suggest that although the processing of new images can interfere with the recognition of previously learned images (Potter & Levy, 1969), the ability to comprehend the semantic meaning of a scene occurs in as little as 120 msec, and that this information can be represented either by visual, or verbal description (see also, Intraub, 1999; Potter, 1999).

Along the same lines, Biederman (1972) investigated the effects of context on perceptual scene recognition. In his task, observers were presented with natural images (e.g., a street scene presented for 300, 500, or 700 msec) and were asked to identity the presence of a target object at a cued location (e.g., a bike). In the pre-cued condition, an arrow preceded presentation of the image and identified the location of the target. In the post-cued condition, the cueing arrow was presented after image offset. The scenes were constructed such that they were either coherent, or jumbled (i.e., the image was cut into six equal segments and rearranged, so that its context was ambiguous) and some observers were provided with target foreknowledge, whereas others were not. Biederman reported that target identification accuracy was lower for jumbled scenes compared to coherent scenes, irrespective of cue type, target knowledge, and presentation duration. As such, Biederman concluded that image coherency facilitated object detection by activating a scene schema, corroborating Potter and colleagues results that contextual scene information appears to be available pre-attentively.

1.2 The time course of rapid scene categorization

The investigations undertaken by Potter (1976) and Biederman (1972) were influential in suggesting that scene context is processed rapidly, and without attention. Inspired by these findings, Thorpe, Fize, and Marlot (1996) were the first to estimate the time course of rapid scene categorization. They asked their participants to indicate the presence (or absence) of an animal in briefly presented (20 msec) natural scenes (unmasked), while simultaneously recording electroencephalography (EEG) activity. Thorpe and colleagues reported that although observers were able to respond within 300 msec of stimulus onset, EEG activity began to differ between target absent and target present trials at approximately 120 msec after stimuli onset. This finding is important because it suggests that the information used to categorize the presence of an animal was available prior to this time point. However, because the authors did not mask the images, it is possible that later processing contributed to the ability to differentiate between target present and target absent displays. Nevertheless, subsequent studies showed that rapid scene categorization cannot be sped up by training (Fabre-Thorpe, Delorme, Marlot, & Thorpe, 2001) and extends to basic (e.g., beach and forest), superordinate (e.g., natural and urban), and non-evolutionary important categories (e.g., vehicles) (Greene & Oliva, 2009; Joubert, Rousselet, Fize, & Fabre-Thorpe, 2007; VanRullen & Thorpe, 2002). However, superordinate level categorization has been shown to occur faster than basic level categorization (Loschky & Larson, 2010; Joubert et al., 2007; Rousselet, Joubert, & Fabre-Thorpe, 2005; Larson & Loschky, 2009).

1.3 Does rapid scene categorization require attention?

The rapid nature of scene categorization suggests that the extraction of semantic information occurs automatically and without the need for attentional resources. Thus, much recent research focuses on whether scene perception satisfies the requirements in order to be considered an automatic process. Brown, Gore, and Carr (2002) outlined that in order for a process to be considered automatic in must satisfy the following three conditions: 1) it must be computed rapidly; 2) without intention; and 3) be immune to interference from concurrent processes. Because of the robust finding that scene categorization occurs in as little as 120 msec, it is widely accepted that it meets the first requirement. However, conflicting findings question whether scene perception satisfies the remaining two criteria. In sections 1.3.1 and 1.3.2, I describe these conflicting sets of findings.

1.3.1 Is semantic scene related information extracted without intention?

Inattentional blindness is the perpetual phenomenon in which a fully visible object goes unnoticed when attention is engaged in completing a task. In a typical task (e.g., Mack & Rock, 1998), observers are presented with a masked presentation of a centrally located fixation cross consisting of a long and short line, displayed for less than 200 msec. The task of the observer is to indicate whether the vertical or horizontal line of the cross is longer. On some trials, the cross is presented alone, whereas on others, a critical stimulus (usually a small square) is presented simultaneously in the periphery. For trials on which a critical stimulus is present, observers are asked to indicate whether a stimulus other than the cross is present. Despite the fact that observers are continually probed to report the addition of a second stimulus throughout the experiment (thereby increasing the chances of an erroneous positive response, or false alarm), observers fail to report the critical stimulus, suggesting that without attention significant changes go unnoticed.

Mack and Rock (1998), however, reported that natural scene perception is immune to inattentional blindness. Mack and Rock modified their original design by replacing the critical stimulus with a large image (e.g., an indoor or outdoor scene) on which the central cross was located. In contrast to the typical finding, observers were able to report the semantic meaning of the scene when probed. Although several researchers (Cohen, Alvarez, & Nakayama, 2013; Lamme, 2003) have cited this finding as evidence that the extraction of semantic information does not require attentional resources, Mack and Rock attributed this result to the fact that the large image size attracted attention. In support of this hypothesis, Mack and Clarke (2012) showed that when the critical natural image was both smaller and presented in the periphery, observers failed to notice its presence, corroborating Mack and Rock's original claim.

Closely related to Mack and Rock's (1998) finding is the observation that scene perception is immune to change blindness; the perceptual phenomenon in which an observer fails to notice a significant change to a stimulus when visual attention is diverted elsewhere (Simons & Chabris, 1999). One way in which change blindness is typically studied is using the flicker paradigm. This paradigm involves the continuous alternation between two images, the second image being a copy of the first with either a large, or small change to a critical stimulus. The task of the observer is to identify the change as quickly and as accurately as possible. Typically, changes between the images can go unnoticed for several minutes, suggesting that attention to the object, or area undergoing the change, is needed in order to detect it.

Amid all this research, however, Rensink, O'Regan, and Clark (1997) noted that changes that alter the meaning of a scene are detected faster than changes that do not. In their study, Rensink and coworkers manipulated the image change, such that the object removed was either central, or marginal to the scene's understanding. When a central interest object was removed, change detection rates were statistically significantly faster than when a marginal object was removed. Simons and Levin (1997) cited this result as evidence that the automatic and pre-attentive extraction of semantic meaning guides attention to objects that are central to a scene's meaning. Marginal changes thus go unnoticed because they do not contribute to overall understanding of the scene.

Further evidence that scene categorization occurs automatically is provided by Greene and Fei-Fei (2014), who used a modified Stroop paradigm (Stroop, 1935). The

Stroop paradigm involves presenting observers with coloured words that are either congruent (e.g., “blue” printed in blue) or incongruent (e.g., “blue” printed in red). The classical finding is that naming the colour of a colour word printed in an incongruent ink colour is more difficult than naming the colour of a colour word printed in a congruent ink colour because in the former case, the output of two automatic processes (i.e., colour naming and word reading) yields conflicting results, whereas no such conflict is present (or even possible) in the latter. The Stroop paradigm is thus an effective task for addressing the question of whether a particular process can be considered automatic. Greene and Fei-Fei presented their observers with either an object (e.g., a guitar) or scene (e.g., a lobby) word that was superimposed onto images that were either congruent (e.g., a picture of a guitar, or lobby), or incongruent (e.g., a picture of a barbeque, or cafeteria). The task of the observer was to as quickly and as accurately as possible identify whether the word corresponded to an object, or scene name. Greene and Fe-Fei reported that discriminating between object and scene names was faster on congruent than incongruent trials, replicating the standard stroop effect. They interpreted this finding as suggesting that the automatic extraction of a scene’s meaning interfered with the word-naming task. However, this effect was only observed for basic level scene categories (e.g., beach and river); there was no evidence of stroop interference when scene categories were defined based on the superordinate level (e.g., natural and outdoor), suggesting that not all rapid categorizations are automatic and pre-attentively processed.

1.3.2 Is scene perception immune to interference from concurrent processing?

Investigations addressing whether scene perception is immune from concurrent processing have typically focused on examining whether scene perception performance differs between single and dual-task conditions. In these studies, observers are asked to compete two concurrent tasks: an attentionally demanding primary task and secondary scene categorization task. In theory, the primary task will deplete attentional resources, thereby allowing no resources to be allocated to the scene stimuli. Evidence of pre-attentive scene categorization is thus present if scene categorization performance does not differ between single and dual task conditions. However, and similar to above, these studies have produced mixed results: whereas some studies have documented a cost of dividing attention on scene categorization performance, others have not. In the following section, I review these conflicting sets of findings.

Li, VanRullen, Koch, and Perona (2002) provided the seminal investigation demonstrating that scene categorization is not impaired under dual-task conditions. On each trial in their experiment, observers were simultaneously presented with a central letter discrimination task (e.g., search for an "L" among rotated "T"s; presented until response at the centre of the screen) and a peripheral scene categorization task (e.g., indicate the presence of an animal, or vehicle; presented for 27 msec at a random location in the periphery). On single-task central trials, observers were instructed to respond as quickly and as accurately as possible to only the attentionally demanding central task; on single-task peripheral trials, observers were instructed to respond to only the peripheral scene categorization task. On dual task trials, observers were asked to answer both tasks as quickly and as accurately as possible. Critically, in the latter condition, observers were instructed to respond to the peripheral task before the central task. Results indicated that scene categorization performance did not statistically significantly differ between the single task conditions (i.e., completion of the central or peripheral tasks only) and the dual task condition. Li and colleagues interpreted this finding as evidence that some high-level representations of a visual scene (e.g., semantic information) are available pre-attentively.

In contrast to the claim that scene categorization is attention free, Walker, Stafford, and Davis (2008) demonstrated an associated cost to scene categorization performance when completed concurrently with an attentionally demanding primary task. Walker and colleagues presented their observers with a natural image (170 msec) superimposed with four letters, arranged in square, located centrally. On single task trials, observers were asked to indicate if the image contained an animal (or not). On dual task trials, observers were instructed to first indicate if the four letters contained a vowel before completing the categorization task. Critically, and in contrast to Li and colleagues (2002), the images used in this study contained from one to four distractor objects that were not animals. The authors reported that scene categorization performance was worse on dual task trials than on single task trials, suggesting that attention is required for the rapid categorization of complex visual scenes (i.e., scenes that contain more than one object). Furthermore, this decrease in performance was also present (although attenuated) for trials on which an image contained only one distractor item, suggesting that some attentional resources are required to rapidly categorize simple visual scenes, as used by Li et al. (2002). Taken together, Walker and colleagues concluded that scene categorization is not pre-attentive.

According to Walker and colleagues (2008), the discrepancy between their findings and those reported by Li and colleagues (2002) is due to the fact that their attention task was more attentionally demanding (e.g., the authors reported a 68%

accuracy rate compared to the titrated accuracy rate of 80% reported by Li et al., 2002). Consistent with this interpretation, Cohen and coworkers (2011) showed that scene categorization is susceptible to inattentional blindness if the primary attention task is sufficiently difficult enough to engage attention (c.f. Mack & Rock, 1998). In Experiment 1, Cohen and coworkers modified the inattentional blindness paradigm to include a concurrent task. They asked observers to complete a motion object-tracking task that required the tracking of 4 of 8 discs on a checkered background. The discs moved under either low (track 4 of 8 discs moving at 4.5° per second) or high (track 4 of 8 discs moving at 10.5° per second) attention demands. On critical trials, the checkered background was replaced with a natural scene (e.g., beach, building, highway, mountain, or indoor scene). In a control condition, observers were instructed not to track any discs. Similar to previous inattentional paradigms, observers on critical trials were probed to indicate the presence of any additional stimuli. Although there was no statistically significant difference in scene categorization detection performance between control and dual task conditions when tracking speed was slow, scene categorization detection performance was impaired under dual-task conditions when tracking speed was fast. As such, this finding corroborates Mack and Clarke's (2012) finding that scene perception is susceptible to inattentional blindness under the right conditions.

In a follow-up experiment, Cohen and colleagues (2011) further modified the inattentional blindness paradigm to demonstrate that scene perception in an RSVP stream is impaired when attention is engaged in completing a concurrent task (c.f. Potter, 1976). Cohen and coworkers presented their observers with a stream of masked letters and digits, appearing one at a time at a presentation rate of 100 msec/letter or digit. The task of the observer was to count the number of digits that appeared in the stream. On critical trials, the second-to-last mask was replaced with an image. In a scene category condition, the image corresponded to one of five possible scene categories: a mountain, beach, highway, indoor, or building scene. In an animal/vehicle condition, the image contained either an animal, or a vehicle. Observers were subsequently probed to identify either the scene category, or to indicate whether an animal, or vehicle was present. In a single task condition, observers were instructed to passively view the RSVP stream. Compared to the single task condition, detection rates for both the scene categorization task and the animal/vehicle tasks were lower under dual-task conditions, corroborating the results from their motion tracking experiment. However, Cohen and coworkers also reported that the cost of dividing attention was greater for the animal detection task, suggesting that animal detection might rely on a different mechanism than scene categorization.

1.4 Scene categorization theories

Although scene categorization occurs rapidly, manipulations of attention affect scene categorization performance. A challenge for researchers is to thus establish a scene categorization theory that explains its rapid nature, while also acknowledging a potential role for attention. Although there is still considerable debate regarding the role of attention, there is general agreement that the basis for rapid scene categorization occurs at early stages of visual input. However, researchers disagree with respect to the type of low-level information that subserves rapid scene perception. Whereas Evans and Treisman (2005) suggested that the extraction of summary statistics underpins rapid scene categorization, Oliva and Schyns (1997) suggested that categorization is based on fine and coarse scale information. In the following sections, I elaborate on the evidence for these two respective theories and in section 1.4.3, I discuss how these two theories potentially integrate.

1.4.1 Summary statistics and scene categorization.

Evans and Treisman (2005; see also Chong & Treisman, 2005; Treisman, 2006) suggested that the automatic extraction of a scene's statistical properties underpins rapid scene categorization. This hypothesis is largely based on the existence of an automatic averaging mechanism that extracts statistical properties (e.g., mean, range, and variance) from sets of similar objects. This averaging mechanism appears to be general in its operation, applying to both low-level features (e.g., average size and orientation, Ariely, 2001; Chong & Treisman, 2003, 2005a, Parkes, Lund, Angelucci, & Solomon, 2001) and higher-level properties (e.g., average emotion of a set of faces; Brand, Oriet, & Sykes-Totteham, 2013; de Fockert & Wolfenstein, 2009; Haberman & Whitney, 2007, 2009). Typically, this type of perceptual averaging is studied using a mean discrimination task, or a member identification task. During these tasks, observers are asked to indicate whether a test probe is a set member of a previously displayed set of items (the member identification task), or to judge whether some characteristic of the test probe (e.g., size) corresponds to the average of that characteristic in the previously displayed set (e.g., the set's overall average size of items; the mean discrimination task). The classical finding is that whereas mean discrimination performance is generally very good, member identification performance is directly related to the statistical association between the test probe and the set of items; the closer the test probe is to average characteristic of the set, the greater the likelihood that the test probe will be identified as a set member, regardless of whether it was or not (Ariely, 2001). On the basis of such findings, it is argued that when observers are presented with a set of similar objects, they automatically extract overall statistical properties, while having a poor representation of individual items (Ariely, 2001; Chong & Treisman, 2003, 2005).

Perceptual averaging occurs in as little as 50 msec (Chong & Treisman, 2003), suggesting that it does not require attentional resources (Alvarez & Oliva, 2008; Chong & Treisman, 2003), and is based on information established by early visual processes before conscious awareness of individual objects (Choo & Franconeri, 2010; Corbett & Oriet, 2011; Haberman & Whitney, 2011; but see Jacoby, Kamke, & Mattingley, 2013; Myczek & Simons, 2008). Consequently, Chong and Treisman (2005) proposed that the rapid extraction of statistical information allows for an economical description of a scene, which in turn, provides the basis for rapid scene categorization. The hypothesis is that although statistical information varies from category to category (e.g., the statistical information relating to beaches is different from forests), it is typically consistent within categories (e.g., all beaches have roughly the same statistical information), allowing for rapid categorization without the need for attention to bind features together.

1.4.2 Scene categorization and spatial scale processing.

According to Schyns and Oliva (1994; see also Oliva & Schyns, 1997) rapid scene perception is based on information encoded by the different spatial frequency channels of the visual system. Their hypothesis is largely based on classical findings from the psychophysical literature. Campbell and Robson (1968), for example, showed that detection of square-wave gratings could be predicted by their individual spatial frequencies, suggesting that the visual system encoded the stimuli via different spatial frequency filters. Subsequent studies showed that visual input is initially filtered into between four to six different spatial frequency channels (Ginsburg, 1986; Wilson & Bergen, 1979), and that this filtering precedes stereopsis (Legge & Gu, 1989), motion perception (Morgan, 1992), depth perception (Marshall, Burbeck, Ariely, Rolland, & Maritn, 1996), and saccade programing (Findlay, Brogan, & Wenban-smith, 1993). Thus, Oliva and Schyns (1994; see also, Morrison & Schyns, 2001) suggested that spatial frequency information is the foundation for visual categorization in which the early filtering of visual information precedes the processing of higher-level scene information. More specifically, Schyns and Oliva argued that spatial frequency information could be broadly classified into relatively low and high spatial frequencies. Whereas low spatial frequencies convey information related to an image's global features, high spatial frequencies convey information related to an image's fine details (see Figure 1.1). Attending to fine and global information will facilitate categorization based on each respective source of information (Oliva & Schyns, 1997).

1.4.3 Summary statistics, spatial scale processing, and scene categorization.

Perceptual averaging, spatial scale processing, and rapid scene categorization are mostly investigated independently, with little or no discussion regarding their integration. This is interesting given that it is hypothesized that the extraction of summary statistics contributes to the information found in spatial scales that have been argued to underpin rapid scene categorization (Oliva & Torralba, 2001; Greene & Oliva, 2009). Consider, for example, the Spatial Envelope Theory of scene perception in which rapid scene categorization is based on orthogonal global features that are represented within an image's spatial scales (Oliva & Torralba, 2001). In their original description, Oliva and Torralba (2001) identified an image's ''naturalness'', ''openness'', ''roughness'', ‘‘expansion’’, and ‘‘ruggedness’’ as the original global properties that can be used as the basis for categorization. In support of their hypothesis, Oliva and Torralba created a set of filters based on each global property by computing the Fourier spectra of images that independent observers rated as displaying each respective global property. The Fourier spectrum plots an image’s spatial frequency information along the cardinal orientations, horizontal, vertical, and oblique. Different categories typically convey different spectral information. For example, the Fourier spectrum of a beach scene contains low spatial frequencies along the vertical axis that corresponds to its horizon. Conversely, the Fourier spectrum of a city scene typically has a wide range of spatial frequencies located on the horizontal axis that represents its skyline.

Oliva and Torralba (2001) constructed a computational model that classified images into different scene categories, based on the responses of the different global filters. Results of their simulation showed that global properties could successfully categorize scenes based on superordinate (e.g., natural and manmade) and basic (e.g., beach and forest) levels. For example, scenes that produced a high value on the “naturalness” filter (e.g., a “natural” scene) tended to have lower spatial frequencies on the horizontal axis. Conversely, scenes that produced a low value on the “naturalness” filter (e.g., a “manmade" scene) tended to have more middle and higher spatial frequencies along the horizontal axis. Thus, Oliva and Torralba argued that superordinate level categorization precedes basic level categorization because it can be accomplished on the basis of a single global feature. In contrast, basic level categorization takes longer because it requires the integration of several global features (e.g., a forest scene is both “natural” and “closed”), a conclusion recently corroborated by Greene and Oliva (2009) and consistent with previous suggestions that superordinate categorization occurs before basic level categorization (Loschky & Larson, 2010; Joubert et al., 2007; Rousselet, Joubert, & Fabre-Thorpe, 2005; Larson & Loschky, 2009).

Although the Fourier amplitude spectrum is able to convey that a natural scene contains low spatial frequencies at a horizontal orientation, it is not able to convey where in the image that information is located (e.g., the middle, top-right corner, left-bottom hand corner, etc...). Information related to the location of an image's spatial frequencies is conveyed by an image's Fourier phase spectrum. Oliva and Torralba (2001) showed that when an image's Fourier phase is randomized (and only contributes unlocalized spatial frequency information related to a scene) their model is able to accomplish scene categorization at an 85% accuracy rate. This is compared to the 92% accuracy rate when an image's Fourier phase was unchanged (i.e., the Fourier phase spectrum conveys localized spatial frequency information), suggesting that the most useful spatial frequency information was unlocalized. However, subsequent behavioural studies showed that randomizing an image's phase impairs both basic (Loschky et al., 2007) and superordinate level categorization (Loschky & Larson, 2008), suggesting a discrepancy between simulation and behavioural data.

1.5 The present studies

As evident in the above discussion, current research is focused on whether attention is needed in order to extract semantic meaning. These studies, however, have led to conflicting results: evidence for pre-attentive scene perception on the one hand, and evidence of impaired scene perception without attention on the other. Together, these conflicting sets of findings suggest that attention is required for rapid scene categorization; however, the exact role of attention is still largely unknown. Thus, the purpose of the present set of studies is to investigate potential roles for attention. This investigation is undertaken by addressing the hypothesis that one of the roles of attention is to facilitate scene categorization. As mentioned in the introduction to this chapter, this investigation is broken down into two main chapters, with each chapter focusing on examining the effects of a different type of attention on scene categorization ability. In the following sections, I describe the rationale for the studies reported in Chapters 2 and 3, but first describe the different types of attention that will be examined in this thesis.

The terms global and local attention and distributed and focused attention are often used interchangeably in the scene perception literature. However, there are key differences between these types of attention that makes this comparison unwarranted. First, the terms global and local refers to how attention may be deployed in a hierarchical fashion. For example, for any natural scene, there is both a global and local structure and it is possible to attend to each level; each attended level yielding a different type of information (as described in the opening section of this introduction). Conversely, distributed and focused attention refers to how attention may be allocated over groups, or individual objects, without regard to a hierarchy. Similar to global and local attention, focused and distributed attention is associated with a different type of scene information (as described in the summary statistics section above). Whereas local and focused types of attention are associated with fine, detailed information, global and distributed types of attention are associated with the ability to rapidly categorize a briefly presented scene. As such, the present work is a first attempt to investigate how these different types of attention affect scene categorization ability. The effects of global and local attention on rapid scene categorization are investigated in Chapter 2. The effects of distributed and focused attention on rapid scene categorization are investigated in Chapter 3.

1.5.1 Chapter 2

The studies in Chapter 2 tested Oliva and Schyns (1997) hypothesis that attending to local and global levels facilitates rapid scene categorization based on fine and coarse information, respectively. The experiments used hybrid images that contain a low spatial frequency version of one image (e.g., a city scene) and a high spatial frequency version of another (e.g., a highway scene). The low spatial frequency image conveys information relating to a hybrid's coarse content whereas the high spatial frequency image conveys information relating to a hybrid's fine content (Schyns & Oliva, 1994). Therefore, hybrid images are ideal for investing spatial scale preference because basing categorization on fine and coarse information, respectively, results in different answers. Although hybrid images have been used to examine spatial scale preference (Schyns & Oliva, 1994; Oliva & Schyns, 1997), there is nevertheless a lack of empirical evidence demonstrating the preferred spatial scale, irrespective of any experimental manipulation, used for scene categorization. For example, although low spatial frequencies are suggested to be the default spatial scale used for scene categorization (Schyns & Oliva, 1994), high spatial frequencies can also be used to classify a natural scene (Oliva & Schyns, 1997) Thus, Experiments 1 and 2 addressed the following question: when given the choice between fine and coarse information, what information does an observer prefer to use as the basis for rapid scene categorization? The results from these two experiments provided the baseline for Experiments 3 and 4, which investigated how attending locally and globally affected the subsequent selection of spatial scale scene information used to categorize hybrid images. Attention to local and global levels was manipulated by asking observers to complete global and local Navon tasks (Navon, 1972). The processing of local and global Navon stimuli is analogous to the processing of fine and coarse information within a scene (Badcock, Whitworth, Badcock, & Lovegrove, 1990; Shulman & Wilson, 1987). Thus, Navon stimuli are the ideal stimuli with which to prime observers to select fine and coarse information in subsequently presented hybrid images.

1.5.2 Chapter 3

The four experiments in Chapter 3 were designed to examine Evans and Treisman's (2005) suggestion that the extraction of summary statistics provides the foundation for rapid scene categorization. To date, research related to this question has been limited to two types of investigations: 1) studies that examined the effects of distributed and focused attention on the extraction of summary statistics (Chong & Treisman, 2005); and 2) inferences made from the results of studies designed to address other issues (e.g., the effects of dividing attention on scene categorization performance; Robitille & Harris, 2011). The experiments reported in Chapter 3 adopted a novel approach to answering this question. They examined whether manipulations of attention known to affect the extraction of summary statistics also affect rapid scene categorization in the same fashion.

The purpose of Experiment 1 was two-fold: 1) to replicate the finding that distributed, rather than focused attention, facilitates the extraction of summary statistics; and 2) to index the effects of perceptual averaging using reaction time so that the results could be easily compared with Experiments 2 – 4 that used the same measure to index scene categorization performance. Experiment 2 addressed the question of whether distributed or focused attention was compatible with the detection of an animal. The decision to use an animal detection task was based on previous research that used this task to measure scene categorization ability (Li et al., 2002; Rousselet et al., 2002; Thorpe et al., 1996; Walker et al., 2008). Whereas early scene categorization studies used animal detection tasks, more recent investigations used natural scenes in order to determine if scene categorization behaviour differed between simple and complex stimuli. Specifically, the majority of ongoing research asks observers to classify scenes based on either their basic (e.g., beach or forest) or superordinate (e.g., natural or manmade) level. The robust finding is that superordinate level categorization is faster than basic level categorization (Loschky & Larson, 2010; Larson & Loschky, 2009; Joubert et al., 2007; Rousselet, Joubert, & Fabre-Thorpe, 2005), suggesting that superordinate level information is available earlier than basic level information. In turn, this suggests that information related to superordinate categorization is potentially available pre-attentively and is thus more likely to be influenced by distributed attention. Experiments 3 and 4 directly tested this hypothesis by examining the effects of distributed and focused attention on basic (Experiment 3) and superordinate (Experiment 4) level categorization.

Chapter 2: Attention to local and global levels of hierarchical Navon figures affects rapid scene categorization

Brand, J., Johnson, P. A. (2014). Attention to hierarchical Navon figures affects rapid scene categorization. *Frontiers in Psychology*, 5, 1274. doi: 10.3389/fpsyg.2014.01274

2.1 Abstract

In four experiments, we investigated how attention to local and global levels of hierarchical Navon figures affected the selection of spatial scale information used in scene categorization. We explored this issue by asking observers to classify hybrid images (i.e., images that contain low spatial frequency content of one image, and high spatial frequency content from a second image) immediately following global and local Navon tasks. Hybrid images can be classified according to either their low, or high spatial frequency content; thus, making them ideal for investigating spatial scale preference. Although observers were sensitive to both spatial scales (Experiment 1), they overwhelmingly preferred to classify hybrids based on low spatial frequency content (Experiment 2). In Experiment 3, we demonstrated that low spatial frequency based hybrid categorization was faster following global Navon tasks, suggesting that low spatial frequency processing associated with global Navon tasks primed the selection of low spatial frequencies in hybrid images. In Experiment 4, replicating Experiment 3 but suppressing the low spatial frequency information in Navon letters by contrast balancing the stimuli examined this hypothesis. Similar to Experiment 3, observers preferred to classify hybrids based on low spatial frequency content; however and in contrast, low spatial frequency based hybrid categorization was slower following global than local Navon tasks.

2.2 Introduction

The ability to perceive a scene under increased attentional load is often cited as evidence of pre-attentive scene perception. This evidence is typically indexed using dual-task paradigms in which a secondary scene categorization task is unaffected by a concurrent, cognitively demanding primary task. Researchers argue that scene perception is pre-attentive as it is immune to inattentional blindness (Mack & Rock, 1998), unimpaired under dual task conditions (Li, VanRullen, Koch, & Perona, 2002; Rousselet, Fabre-Thorpe, & Thorpe, 2002), susceptible to stroop interference (Greene & Fei-Fei, 2014), and impervious to change blindness if the object's removal does not change the meaning of the scene (Rensink, O'Regan, & Clark, 1997; Simons & Levin, 1997).

However, other researchers question the evidence in support of the automaticity of scene perception. Cohen, Alvarez, and Nakayama (2011) argued that previous studies falsely demonstrated pre-attentive scene perception because they failed to use sufficiently demanding primary tasks, thereby allowing attentional resources to be allocated to the scene stimuli. By increasing the primary task difficulty, Cohen and colleagues demonstrated that concurrently completing multiple-object tracking and serial representation visual presentation (RSVP) tasks impairs scene categorization. Together with previous research in which deficits in scene perception were indexed using attentional blink (Marois, Yi, & Chun, 2004; Slagter, Johnstone, Beets, & Davidson, 2010; Evans & Treisman, 2005), inattentional blindness (Mack & Clarke, 2012), and dual task (Walker, Stafford, & Davis, 2008) paradigms, Cohen and colleagues concluded that conscious scene perception requires attention.

Although concluding that attention is necessary for a scene to reach conscious awareness, Cohen and colleagues (2011) acknowledged that some higher-level aspects of scene processing occur in the absence of attention. One of the strongest findings in support of this hypothesis is the presence of scene-related behaviours that occur so rapid that attention is thought to play little or no role. Kirchner and Thorpe's (2006; see also, Crouzet, Kirchner, & Thorpe, 2010) study illustrates this point. They showed that when two natural images are presented concurrently, observers are able to make an ultra-rapid saccade to the image that contained an animal in as little as 120-130 msec. Consistent with this view, Thorpe, Fize, and Marlot (1996; see also, Fabre-Thorpe, Delorme, Marlot, & Thorpe, 2001) showed that observers are able to remove their finger from a button box within 300 msec in response to the presence of an animal. Critically, simultaneous event-related potentials revealed a differential frontal lobe activity between target and non-target displays approximately 150 msec after stimulus onset. This suggests that scene categorization is made prior to this time point. Researchers (VanRullen & Thorpe, 2002) cite such results as evidence that scene categorization is accomplished, in part, by an automatic feed-forward mechanism, a conclusion corroborated by simulation evidence (Serre, Oliva, & Poggio, 2007).

The rapid ability to categorize scenes suggests that a scene's semantic content is based on information originating from early visual processes. Consistent with this idea, Schyns and Oliva (1994) suggested that rapid scene categorization is based on a scene's global layout. Highways, for example, tend to have fewer vertical straight lines compared to city landscapes that have many dense, vertical orientations. Although these global image properties can vary from one scene to another (e.g., some cities are less dense than others), the consistency of spatial organization across different scenes is thought to activate a scene schema that can be used for rapid scene categorization. Schyns and Oliva tested this hypothesis by introducing a new type of scene stimuli, termed a hybrid image. A hybrid image contains information from two separate sources at different spatial frequencies. For example, an image that contains the low spatial frequency content of one picture (e.g., a city scene), and the high spatial frequency content of a second picture (e.g., a highway scene). Of particular importance to Schyns and Oliva was not spatial frequency per se, but rather the information that each spatial scale conveyed for scene recognition. Converging evidence from neurophysiological and psychophysical studies suggest that visual information is organized into spatial frequency channels in which global information is conveyed by low-spatial frequencies (LSFs) and finer information is conveyed by high-spatial frequencies (HSFs; for a review, see Morrison & Schyns, 2001). Consequently, the authors reasoned that if scene recognition is based on coarse information, then observers should prefer to categorize hybrid images based on LSF content.

To test their hypothesis, Schyns and Oliva (1994) asked observers to indicate whether a briefly presented (30 msec or 150 msec) sample image matched a subsequent target image. The sample image was either a hybrid, low-pass filtered (i.e., contained only LSFs), high-pass filtered (i.e., contained only HSFs), or a full broadband spatial frequency scene (i.e., an unaltered original image). The target image was always a broadband image. Of critical importance here was the association between hybrid samples and target images. On LSF-hybrid trials, the hybrid's LSF content matched the target scene. On HSF-hybrid trials, the hybrid's HSF content matched the target scene. When presentation duration was short, LSF-hybrid trials were more accurate than HSF-hybrid trials; conversely, when presentation duration was long, HSF-hybrid trials were more accurate than LSF-hybrid trials. Critically, categorization performance was high for all control conditions, suggesting that differences in spatial frequency availability cannot account for the differential processing of hybrid images. Schyns and Oliva attributed this result to a coarse-to-fine processing bias in which the early availability of a scene's global layout activates a scene schema from memory. Finer details emerge later and fill in the details of the scene's content (e.g., object recognition).

Oliva and Schyns (1997) modified the coarse-to-fine hypothesis to reflect the fact that either global, or fine scale information can be used for scene recognition. They asked observers to first complete a sensitization phase during which they were briefly presented natural images that were meaningful at only one spatial frequency (e.g., a LSF version of a highway scene with HSF structured noise). A test phase immediately followed in which observers were asked to classify hybrid images. Observers were more likely to categorize hybrids based on LSF and HSF content, respectively, if they were first sensitized to the same frequencies during the sensitization phase. Interestingly, observers claimed to be aware of only a single spatial scale within the hybrid images, suggesting that scale selection was based on the scale that was previously the most informative.

To explain this flexibility in spatial scale selection, Oliva and Schyns (1997) suggested that attention is driven to spatial frequencies in which recognition is based on scale specific cues of a scene category (e.g., natural landscapes contain low spatial frequencies at a horizontal orientation that correspond to the horizon). This idea dovetails with Chong and Treisman's (2005) notion that different distributions of attention facilitate the extraction of different types of information within a scene. According to

Chong and Treisman, a scene's layout is organized hierarchically and attention can be deployed either locally, globally, or distributed over a set of similar items. When attention is focused locally, features are bound together resulting in the identification of an object. In contrast, when attention is distributed globally, the gist or semantic meaning of a scene is extracted based on its global layout. Finally, when attention is distributed over a set of similar items, summary representations of set properties are automatically extracted (e.g., average size; Ariely, 2001).

Global and local distributions of attention are typically studied using hierarchical Navon stimuli (e.g., a large "A" comprised of smaller "Cs"). Navon (1977) reported a global precedence effect that is characterized by two robust findings. First, global letters are identified faster than local letters; and second, global recognition interferes with local recognition but not vice versa. Several researchers (Badcock, Whitworth, Badcock, & Lovegrove, 1990; Shulman & Wilson, 1987) explained the global precedence effect using the coarse-to-fine processing framework. Similar to the identification of coarse and fine information, the hypothesis is that the identification of global and local information is based on LSF and HSF information, respectively. In addition, Flevaris, Bentin, and Robertson (2011) showed that adopting different attentional distributions facilitates the selection of different spatial scales. They asked participants to classify the orientation of either the LSF or HSF component of a compound sine-wave grating immediately following global, or local Navon tasks. When discriminating the orientation of the LSF component, observers were faster following global Navon tasks; conversely, when asked to discriminate the orientation of the HSF component, observers were faster following local Navon tasks.

Flevaris and coworkers' (2011) result suggests that attending to global and local levels should differentially affect scene categorization by facilitating the selection of LSFs and HSFs, respectively. In the present research, we tested this hypothesis by asking participants to categorize briefly presented hybrid images following global, or local Navon tasks. However, because hybrid images contain competing sources of categorization content, it was important that we first demonstrated the ability of our observers to extract both sources of information. Additionally, it was also important that we understood the spatial frequency that our observers preferred to use for categorization, irrespective of any attention manipulation. Thus, in Experiment 1 we assessed spatial scale sensitivity and in Experiment 2 we assessed spatial scale preference.

Experiment 1 was a probe design similar to Schyns and Oliva (1994) in which observers were asked to indicate whether a probe word matched a briefly presented (32 msec or 150 msec) hybrid image. The probe word matched either the hybrid's LSF, or HSF content. In a control condition, the probe word matched neither spatial frequency. The measure, *d* prime (*d'*) was computed to measure observers' sensitivity to both LSFs, and HSFs. d' values were above 1.5 in each condition, suggesting that both LSFs and HSFs are available in our hybrid images, at both short and long durations. Experiment 2 was a replication of Experiment 1, with the exception that we used an all-alternative forced choice paradigm in which observers were asked to choose the image category from a list of all possible target categories. Critically, this design allowed us to compute an objective measure of preferred spatial scale. Results indicated that observers preferred to categorize hybrid images based on LSF content, at both short and long durations. Together with the results of Experiment 1, Experiment 2 demonstrated that our observers preferred to base categorization on LSF content, despite the fact that both LSFs, and HSFs were perceptually available.

The fact that our observers prefered to base hybrid categorization on LSF content suggests that attending globally facilitates scene categorization. A consequence of this prediction is that LSF-based hybrid categorization should be faster following global compared to local Navon tasks. In Experiment 3, we directly tested this hypothesis by asking observers to classify hybrid images immediately following global and local Navon tasks. Similar to Experiment 2, observers preferred to categorize hybrid images based on LSF content, following both local and global Navon tasks. Furthermore, and consistent with our hypothesis, LSF-based hybrid image categorization was faster following global Navon tasks. In Experiment 4, we directly tested whether this facilitation effect was the result of processing LSFs associated with a Navon figure's global structure. We thus replicated Experiment 3 with the exception that we contrast balanced the Navon stimuli in order to suppress their LSFs (see Appendix 2.1). Similar to Experiment 3, observers preferred to classify hybrid images based on LSF content, irrespective of the Navon task completed; however, and in contrast to Experiment 3, LSF-based hybrid image categorization was slower following global than local Navon tasks.

2.3 Experiment 1

The goal of Experiment 1 was to demonstrate the availability of both spatial frequencies in our hybrid images. Similar to Schyns and Oliva's (1994) study, we asked observers to complete a classification task in which they were required to indicate whether a cue word corresponded to a previously presented low-pass, high-pass, broadband, or hybrid image.

2.3.1 Method

2.3.1.1 Observers. Eight undergraduate students from Concordia University participated in this study in return for partial course credit. All observers self-reported normal or corrected-to-normal vision. The University Human Research Ethics Committee at Concordia University approved all experiments reported in this article and all observers provided written consent.

2.3.1.2 Stimuli and apparatus. Stimuli were presented on a 21-in. Viewsonic 225fb CRT monitor (1024 X 768 resolution; 100 Hz refresh rate) controlled by a Dell Precision T3400 core2 quad processor running Microsoft Windows 7. Experiment Builder (SR Research, Ottawa, Ontario) was used to display the stimuli and record the responses. All participants were seated 60 cm away from the screen, and their head position was controlled using a table-mounted chinrest.

Stimuli were 128 natural images (32 unique images of highways, cities, living rooms, and valleys, respectively) taken from the Sun image database (Zoo, Xiao, Hays, Ehinger, Oliva, & Torralba, 2010). All images were equalized for mean luminance and RMS contrast (as described in Appendix B of Loschky et al., 2007) and were presented on a gray background (RBG values = [128 128 128]; luminance of 52 cd/m^2). These images were the same categories used by Schyns and Oliva (1994), who showed that their overall contrast was similar (i.e., the Fourier amplitude spectra of the images are highly correlated with one another). Images were broadband, low-pass (below 2 cycles deg^{-1} of visual angle), high-pass (above 6 cycles deg^{-1} of visual angle), or hybrid images. Hybrids were constructed by combining the low frequency components of one scene (e.g., a city) with the high frequency components of another scene (e.g., highway). Mathwork Matlab (ver. 2011b) was used to create the images. A total of 32,768 possible hybrid images were constructed by taking every possible combination of the four scene categories. All images were gray scaled, located in the centre of the screen, and were 256 X 256 pixels.

2.3.1.3 Procedure. A trial schematic is presented in Figure 2.1. Each trial began with a fixation cross located in the centre of the screen presented for 250 msec, followed by a single image presented for either 32, or 150 msec. A white noise mask (amplitude spectrum slope = 0; orientation magnitude = 0) immediately followed offset of the image and was presented for 64 msec. The image was a broadband, low-pass, high-pass, or a hybrid image. Immediately following offset of the mask, observers were presented with a display screen in which they were asked to indicate whether a probe word (e.g., highway, city, living room, or valley) corresponded to the category of the previously presented image. On 50% of trials, the cue word corresponded to the image category. Of those 50% of trials on which the image was a hybrid, the probe word matched the hybrid's LSF and HSF content 25% of the time, respectively. We instructed observers to press "1" on the keyboard number pad if they believed the probe word matched the previously presented image and the "2" key if they believed that it did not. The probe word was displayed in the centre of the screen and stayed visible until a response was made. Trial-to-trial feedback was not provided.

Observers completed 16 blocks of 48 trials for a total of 768 trials. Image type and presentation duration varied from trial-to-trial within a block, and the order of images and presentation duration was chosen at random by the program. Observers completed 32 practice trials prior to beginning the experiment. The scene categories used during the practice trials were not used in the experimental trials (e.g., forest and barn scenes) and practice trials were not analyzed.

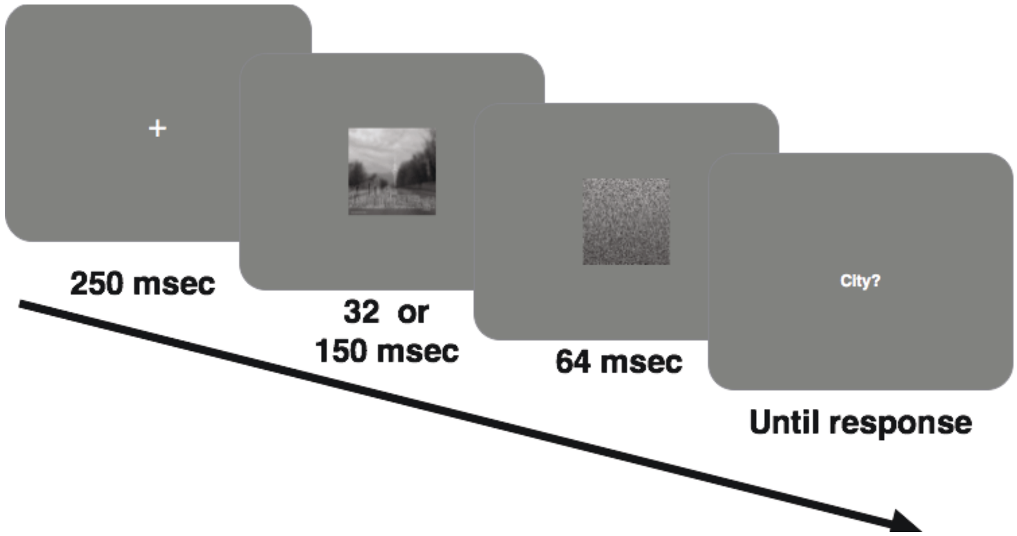


Figure 2.1. An example of a trial sequence in Experiment 1.

2.3.2 Results

2.3.2.1 Sensitivity. The sensitivity measure, *d*' was calculated for each condition. Condition varied according to image type (broadband, low-pass, high-pass, and hybrid) and presentation duration (32 msec and 150 msec). Because hybrid images contained both low and HSF content, we further separated these trials into those on which the probe word matched the hybrid's low (Hybrid-LSF) and HSF content (Hybrid-HSF). As can be seen in Figure 2.2A, *d*' values were high (*d*' > 1.5) in all conditions, suggesting that observers were sensitive to all image types at both presentation durations. We entered *d'* values into a 5 (image type) X 2 (presentation duration) repeated measures Analysis of Variance (ANOVA). There were significant main effects of image type, $F(4, 28) = 8.09$, $p < .001$, $\eta^2 = .54$, and presentation duration, $F(1, 7) = 34.47$, $p < .001$, $\eta^2 = .83$. The image type X presentation duration interaction was also significant, $F(4, 28) = 4.65$, $p < .001$, $\eta^2 = .39$.

Because experiment 1 was designed to determine the availability of spatial frequencies in our hybrid images, we were particularly interested in comparisons between Hybrid-LSF and Hybrid-HSF trial types. However, we first compared performance between control images (low-pass, high-pass, and broadband) in order to ensure that any observed difference between hybrid trial types cannot be accounted for by processing biases in spatial frequency information. We first computed the planned comparison comparing *d*' values using a 3 (image type) X 2 (presentation duration) planned contrast. This contrast was not significant, suggesting that there was no statistical difference in spatial frequency processing as a function of presentation duration, $F(1, 7) = 1.38$, $p > .279$, $\eta^2 < .01$. We then compared sensitivity between control images using a series of contrast comparisons. Specifically, we computed contrasts comparing *d*' values between broadband images and high-pass (Ψ_1) and low-pass (Ψ_2) filtered images, respectively. *d′* statistics and the results of these contrasts are displayed in Table 2.1. Observers were more sensitive to broadband images (M = 3.55; SD = 0.45) than both high-pass (M = 2.88; SD = 0.62) and low-pass filtered images (M = 2.33; SD = 0.40). Observers were equally sensitive to low-pass and high-pass filtered images (Ψ_3). The effect size measures in Experiment 1 paralleled the significance results. The largest effect sizes were between broadband images and low-pass (η^2 = .76) and high-pass (η^2 = .47) filtered images. The effect size between low-pass and high-pass filtered images was relatively smaller in comparison (η^2 = .26).

Following the control image type analysis, we computed the contrast comparing hybrid trial types (Hybrid – LSF and Hybrid – HSF) as a function of presentation duration. This was not statistically significant, $F(1, 7) = .137$, $p > .722$, $\eta^2 < .01$. We followed up this analysis by comparing sensitivity between hybrid trial types using a planned contrast, collapsing over presentation duration (Ψ_4). Observers were more sensitive to hybrid-HSF image types (M = 2.71; SD = 0.49) than hybrid-LSF image types (M = 2.14; SD = 0.26). Furthermore, the associated effect size (η^2 = .66) was similar to the effect sizes reported for the significant control contrasts, suggesting that observers were in fact more sensitive to HSFs than LSFs in the hybrid images.

Table 2.1. *d prime* statistics for each image type at each presentation duration in Experiment 1. *d prime* mean difference contrasts in Experiment 1

	d'					
	32 msec			**150 msec**		
Trial Type	***M***	***SD***	**95% CI**	***M***	***SD***	**95% CI**
Broadband	3.32	0.46	[2.94, 3.71]	3.77	0.58	[3.28, 4.26]
Low-pass	1.93	0.45	[1.56, 2.31]	2.74	0.59	[2.24, 3.23]
High-pass	2.40	0.75	[1.78, 3.02]	3.37	0.62	[2.85, 3.89]
Hybrid-LSF	2.11	0.26	[1.89, 2.32]	2.18	0.38	[1.86, 2.50]
Hybrid-HSF	2.64	0.52	[2.21, 3.07]	2.77	0.47	[2.37, 3.16]

Contrasts

Contrast	***df***	***F***	***p***	M_D	$SE(M_D)$	**95% CI**(M_D)	η^2
Ψ_1	(1, 7)	6.18	< .042	0.67	0.70	[0.09, 1.25]	.47
Ψ_2	(1, 7)	22.2	< .002	1.22	0.24	[0.65, 1.79]	.76
Ψ_3	(1, 7)	2.50	> .158	0.55	0.33	[-0.22, 1.33]	.26
Ψ_4	(1, 7)	13.7	< .008	0.56	0.14	[0.22, 0.89]	.66

Ψ1: *d'* comparison between broadband images and high-pass filtered images.
Ψ2: *d'* comparison between broadband images and low-pass filtered images.
Ψ3: *d'* comparison between low-pass filtered and high-passed filtered images.
Ψ4: *d'* comparison between Hybrid – LSF and Hybrid – HSF image types.

2.3.2.2 Reaction time. We calculated mean reaction time (RT) measures for each trial type as a function of presentation duration. These means are displayed in Figure 2.2B. We entered these means into a 4 (image type) X 2 (presentation duration) repeated measures ANOVA. Unlike the calculation of *d*' statistics, hybrid images were not separated further because target absent trials are the same between Hybrid – LSF and Hybrid – HSF trial types. The main effect of image type was significant, $F(3, 21) = 3.29$, $p < .04$, $\eta^2 = .32$. However, the main effect of presentation duration and the image type X presentation duration interaction were not: $F(1, 7) = .368$, $p > .563$, $\eta^2 < .05$ and $F(3, 21) = .009$, $> .899$, $\eta^2 < .001$.

Similar to the sensitivity analysis, we were primarily interested in differences between Hybrid-HSF and Hybrid-LSF image types, but first report the results related to the control images. Specifically, we computed contrasts that paralleled the sensitivity comparisons. Reaction time statistics and mean difference contrasts are displayed in Table 2.2. Observers were faster to respond to broadband images ($M = 950.04$; $SD = 58.18$) than both high-pass ($M = 1005.26$; $SD = 36.67$) (Ψ_1) and low-pass filtered images ($M = 1007.03$; $SD = 48.75$) (Ψ_2). There was no RT difference between low-pass and high-pass filtered images (Ψ_3). Consistent with the sensitivity analysis, the largest effect size was between broadband images and low-pass filtered images ($\eta^2 = .52$) followed by the effect size for the difference between broadband images and high-pass filtered images ($\eta^2 = .38$). The effect size between low-pass and high-pass filtered images was negligible ($\eta^2 < .01$).

Reaction times on target present trials were compared between Hybrid – LSF and Hybrid – HSF image types and are displayed in Figure 2.2C. We entered these means into a 2 (hybrid trial) X 2 (presentation duration) planned contrast. Consistent with the sensitivity analysis, this contrast was not significant, suggesting that RTs did not differ between hybrid image types as a function of presentation duration, $F(1, 7) = .617$, $p > .458$, $\eta^2 < .08$. We then compared RTs between hybrid – LSF and hybrid – HSF image types, collapsing over presentation duration. This contrast was significant, $F(1, 7) = 7.58$, $p < .028$, $\eta^2 = .52$. Observers were faster to respond to Hybrid – LSF image types (M = 1013.81; SD = 16.37) than Hybrid – HSF image types (M = 1068.62; SD = 41.90). This was a difference of approximately 54.81 msec (SD =52.65; 95% CI [11.91, 97.71]). It is interesting to note that the associated effect size was similar to the effect size reported in the parallel sensitivity analysis ($\eta^2 = .66$), suggesting that the effect of hybrid trial type is robust across dependent variables.

Table 2.2. Reaction time statistics for each image type at each presentation duration in Experiment 1. Reaction time mean difference contrasts in Experiment 1.

	Reaction Time (msec)					
	32 msec			150 msec		
Trial Type	***M***	***SD***	**95% CI**	***M***	***SD***	**95% CI**
Broadband	945.56	78.42	[879.99, 1011.13]	954.51	76.04	[890.92, 1018.09]
Low-pass	1000.89	76.27	[937.11, 1064.67]	1013.17	77.22	[948.59, 1077.73]
High-pass	999.98	46.74	[960.89, 1039.07]	1010.54	52.44	[966.69, 1054.39]
Hybrid	1020.98	52.17	[977.35, 1064.61]	1061.45	41.90	[1026.42, 1096.49]
***Hybrid - LSF**	1001.99	24.80	[981.25, 1022.73]	1025.64	14.56	[1013.46, 1037.82]
***Hybrid – HSF**	1039.97	92.13	[962.93, 1117.00]	1097.27	75.02	[1034.54, 1160.01]

Contrasts

Contrast	df	F	p	M_D	$SE(M_D)$	**95% CI** (M_D)	η^2
Ψ_1	(1, 7)	4.32	< .050	55.22	22.84	[1.21, 109.24]	.38
Ψ_2	(1, 7)	7.58	< .028	56.99	19.37	[11.18, 102.81]	.52
Ψ_3	(1, 7)	.007	> .937	1.76	20.11	[-45.81, 49.34]	< .01

Ψ1: RT comparison between broadband images and high-pass filtered images.
Ψ2: RT comparison between broadband images and low-pass filtered images.
Ψ3: RT comparison between low-pass filtered and high-passed filtered images
* Reaction time calculation is based on target present trials only.

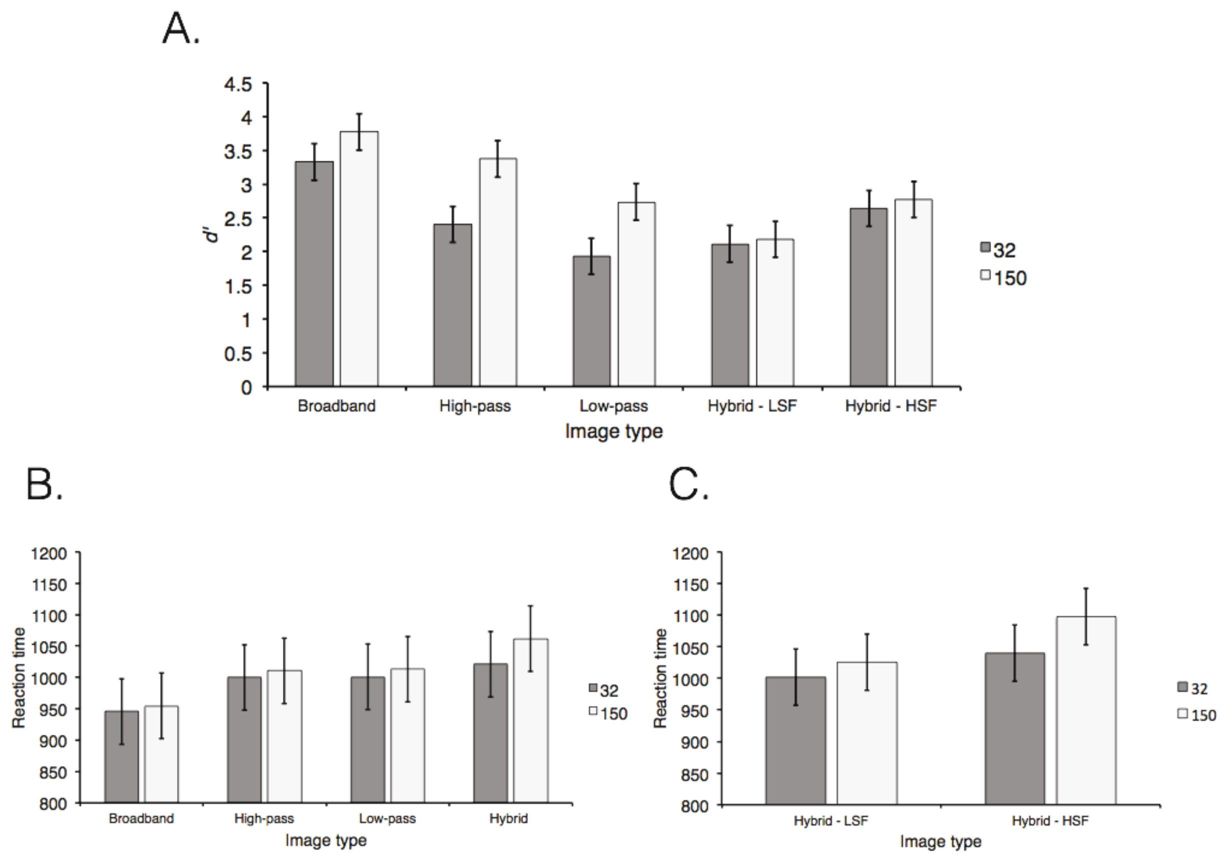


Figure 2.2 The results of Experiment 1. A) Mean *d prime* values for each image type at each presentation duration. B) Mean scene categorization reaction times for each image type at each presentation duration; C) Mean reaction times for hybrid LSF and hybrid- HSF trial types at each presentation duration. The error bars represented here are the 95% within -subject confidence intervals described by Loftus and Masson (1994).

2.3.3 Discussion

The critical result from Experiment 1 is that we corroborated Oliva and Schyns (1997) finding that both spatial scales are available to form the basis for hybrid image categorization. Observers in our study were sensitive to both sources of spatial frequency content and there was no significant interaction with presentation duration, although observers were overall more sensitive to HSFs than LSFs in the hybrid images. An interesting finding from Experiment 1 is that *d*' values were overall high, which is suggestive of weak masking effects. The most likely explanation for this result is that we constructed our masks so that their amplitude spectrum slope (i.e., the slope that conveys amplitude and orientation information in an image) would have a value of 0. Hansen and Loschky (2013) found that white noise masks with this property are the least effective at masking natural scene stimuli, whereas white noise masks whose amplitude spectrum slope most closely resembled that of a natural scene (e.g., ~ alpha = 1; Hansen, Haun, & Essock, 2008) are the most effective. This suggestion is consistent with previous studies that showed that the most effective mask for a particular spatial frequency is one whose amplitude spectrum information is most similar to the target stimuli (Losada & Mullen, 1995; Mullen & Losada, 1999; Stromeyer & Julsez, 1972).

2.4 Experiment 2

Experiment 2 is an extension of Experiment 1. Whereas Experiment 1 assessed the availability of spatial scale information, Experiment 2 assessed spatial scale preference between competing sources of LSF and HSF information. Thus, Experiment 2 is a replication of Experiment 1, with the exception that we assessed scene categorization using an all-alternative forced choice paradigm. We asked observers to choose which of all possible target categories corresponded to the previously presented hybrid image. Because a hybrid image's LSFs and HSFs convey information related to different categories, forcing observers to choose between all possible target categories indexes their preferred spatial scale.

2.4.1 Method

2.4.1.1 Observers. Ten undergraduate students from Concordia University participated in this study in return for partial course credit. All observers self-reported normal or corrected-to-normal vision.

2.4.1.2 Stimuli, apparatus, and procedure. An example of a trial sequence in Experiment 2 is presented in Figure 2.3. Stimuli, apparatus, and procedure were the same as in Experiment 1 with the following exception. Categorization performance was measured using a 4-alternative forced choice task. Immediately following offset of the mask, we presented observers with a list of 4 probe words with an associated number (city = 1, highway = 2, living room = 3, and valley = 4) listed vertically in the centre of the screen. The task of the observer was to as quickly and as accurately as possible indicate the category of the previous image by pressing the corresponding key on the keyboard number pad.

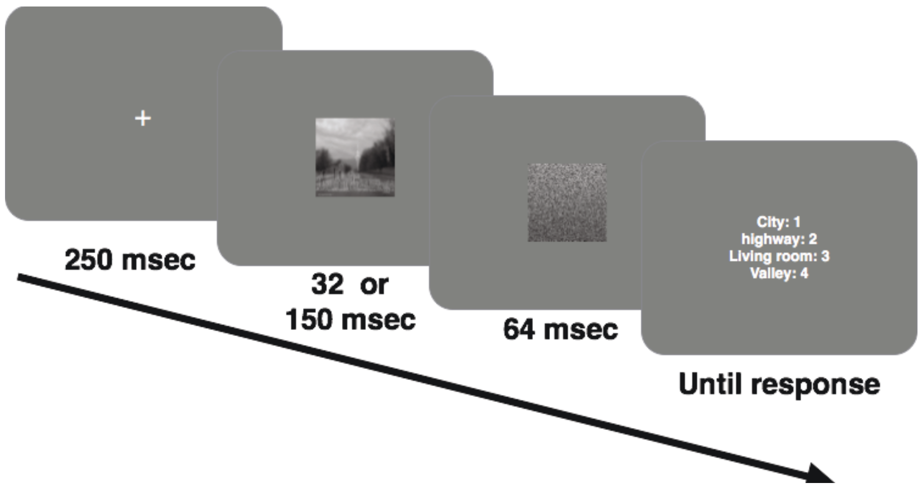


Figure 2.3. An example of a trial sequence in Experiment 2.

2.4.2 Results

2.4.2.1 Sensitivity. As in Experiment 1, *d′* was computed for each condition. These means are displayed in Figure 2.4A. Similar to Experiment 1, *d′* values were above 1.5 in each condition, suggesting that observers were sensitive to all image types. We entered these means into a 2 (presentation duration) X 4 (image type) repeated measures ANOVA. There were significant main effects of image type, $F(3, 27) = 10.91$, $p < .001$, $\eta^2 = .55$, and presentation duration, $F(1, 9) = 56.83$, $p < .001$, $\eta^2 = .86$. The image type X presentation duration interaction was not significant, $F(3, 27) = 1.29$, $p > .299$, $\eta^2 = .13$. Observers were more sensitive at long ($M = 3.05$; $SD = 0.29$) than short ($M = 2.32$; $SD = 0.13$) durations, a difference of 0.73 ($SD = 0.29$; 95% CI [0.52, 0.94]). Although sensitivity was high in all conditions, the significant image type main effect appears to be driven by the fact that observers were less sensitive to hybrid images ($M = 2.04$; $SD = 0.18$) than the other image types ($M = 2.89$; $SD = 0.21$). This contrast (Ψ_1) was statistically significant. Furthermore, the contrast comparing sensitivity between broadband images ($M = 3.09$; $SD = 0.47$) and low-pass and high-pass filtered images ($M = 2.79$; $SD = 0.14$) was not significant, corroborating our conclusion (Ψ_2). Consistent with this conclusion, the effect size for Ψ_1 ($\eta^2 = .85$) was higher than Ψ_2 ($\eta^2 = .31$). *d′* statistics and contrast analyses are displayed in Table 2.3.

In order to examine spatial scale preference, we separated hybrid trials into those on which categorization was based on low and HSF content, respectively. As can be seen in Figure 2.4B, observers preferred to categorize hybrid images based on LSF content at both short and long presentation durations. High spatial frequency based hybrid categorization did not exceed chance at long durations ($M = .20$; $SD = .07$), $t(9) = 1.87$, *p*

> .095 and was worse than chance at short durations (M = .15; SD = .03), $t(9) = 8.64$, $p <$ .001. As a result, we concentrated our analysis on trials on which hybrid categorization was based on LSF content. LSF-based hybrid categorization did not statistically significantly differ between short (M = .73; SD = .08) and long (M = .69; SD = .09) durations, $t(9) = 1.78$, $p > .111$, Cohen's d = .55, a difference of .04 (SD = 0.06; 95% CI [-.01, 0.09]).

Table 2.3. *d prime* statistics for each trial type at each presentation duration in Experiment 2. *d prime* mean difference contrasts in Experiment 2.

	d'					
	32 msec			**150 msec**		
Trial Type	***M***	***SD***	**95% CI**	***M***	***SD***	**95% CI**
Broadband	2.88	0.59	[2.50, 3.26]	3.31	0.38	[3.06 3.55]
Low-pass	2.33	0.59	[1.94, 2.71]	3.13	0.37	[2.89, 3.37]
High-pass	2.39	0.52	[2.06, 2.73]	3.34	0.65	[2.92, 3.75]
Hybrid	1.67	0.28	[1.49, 1.85]	2.41	0.17	[2.31 2.53]

Contrasts

Contrast	***df***	***F***	***p***	M_D	$SE (M_D)$	**95% CI** (M_D)	η^2
Ψ_1	(1, 9)	48.46	< .001	0.85	0.12	[0.39, 1.32]	.85
Ψ_2	(1, 9)	3.94	> .078	0.29	0.14	[-0.28, 0.86]	.31

Ψ1: *d'* comparison between hybrid image types and the other image types.
Ψ2: *d'* comparison between broadband images and low-pass and high pass filtered images.

2.4.2.2 Reaction time. Reaction times were computed as described in the sensitivity analysis and are displayed in Figure 2.4C. We entered RTs into a 2 (presentation duration) X 4 (image type) repeated measures ANOVA. There was a significant main effect of image type, $F(3, 27) = 15.44$, $p < .001$, $\eta^2 = .63$. The main effect of presentation duration and the image type X presentation duration interaction were not significant, $F(1, 9) = .033$, $p > .860$, $\eta^2 = .01$ and $F(1, 9) = 1.77$, $p > .176$, $\eta^2 = .16$. Looking at Figure 2.4C, it is clear that observers were overall slower to respond to hybrid images ($M = 1078.79$; $SD = 131.27$) than any other image type ($M = 838.87$; $SD = 43.76$). This contrast was statistically significant (Ψ_1). Furthermore, observers were faster to respond to broadband images ($M = 795.27$; $SD = 41.91$) than low-pass and high-pass filtered images ($M = 860.66$; $SD = 49.58$) (Ψ_2). There was no significant difference in RTs between low-pass ($M = 890.67$; $SD = 89.6$) and high-pass filtered ($M = 830.65$; $SD = 29.25$) images (Ψ_3). Similar to the previous experiments, effect size comparisons paralleled the significance results. The effect size associated with the non-significant difference between low-pass and high-pass filtered images was the smallest ($\eta^2 = .31$), whereas the largest effect sizes were between broadband images and low-pass and high pass filtered images ($\eta^2 = .75$) and between hybrid images and the other image types ($\eta^2 = .72$). Reaction time statistics and contrast analyses are displayed in Table 2.4.

As with the sensitivity analysis, our main goal was to index differences relating to hybrid images. However, because HSF-based categorization was no better (or worse) than chance, we restricted our hybrid RT analysis to trials on which hybrid categorization was based on LSF content (Figure 2.4D). LSF-based hybrid categorization was statistically significantly faster at short than long durations, $t(9) = 2.98$, $p < .016$, Cohen's $d = .94$, a difference of 108.21 msec ($SD = 109.12$; 95% CI [30.15, 186.28]).

Table 2.4. Reaction time statistics for each trial type at each presentation duration in Experiment 2. Reaction time mean difference contrasts in Experiment 2.

	Reaction time (msec)					
	32 msec			150 msec		
Trial Type	***M***	***SD***	**95% CI**	***M***	***SD***	**95% CI**
Broadband	815.299	104.75	[740.37, 890.23]	775.25	101.48	[702.66, 847.84]
Low-pass	889.69	108.20	[812.44, 966.94]	891.66	115.17	[809.27,974.04]
High-pass	844.03	93.96	[776.81, 911.24]	817.28	82.28	[758.42, 876.13]
Hybrid	1029.46	141.24	[928.42, 1130.49]	1128.08	148.12	[1022.13, 1234.03]
Hybrid - LSF	896.45	157.08	[739.36, 1053.54]	1004.67	144.88	[859.78, 1149.55]

Contrasts

Contrast	***df***	***F***	***p***	M_D	$SE(M_D)$	**95% CI**(M_D)	η^2
Ψ_1	(1, 9)	23.64	<. 001	239.92	55.35	[114.69, 365.11]).	.72
Ψ_2	(1, 9)	27.65	<. 001	65.38	11.80	[38.70, 92.08]	.75
Ψ_3	(1, 9)	4.09	> .074	60.02	28.17	[-3.07, 123.75]	.31

Ψ1: RT comparison between hybrid images and the other image types.
Ψ2: RT comparison between broadband images and high-pass and low-pass filtered images.
Ψ3: RT comparison between high-pass and low pass filtered images.

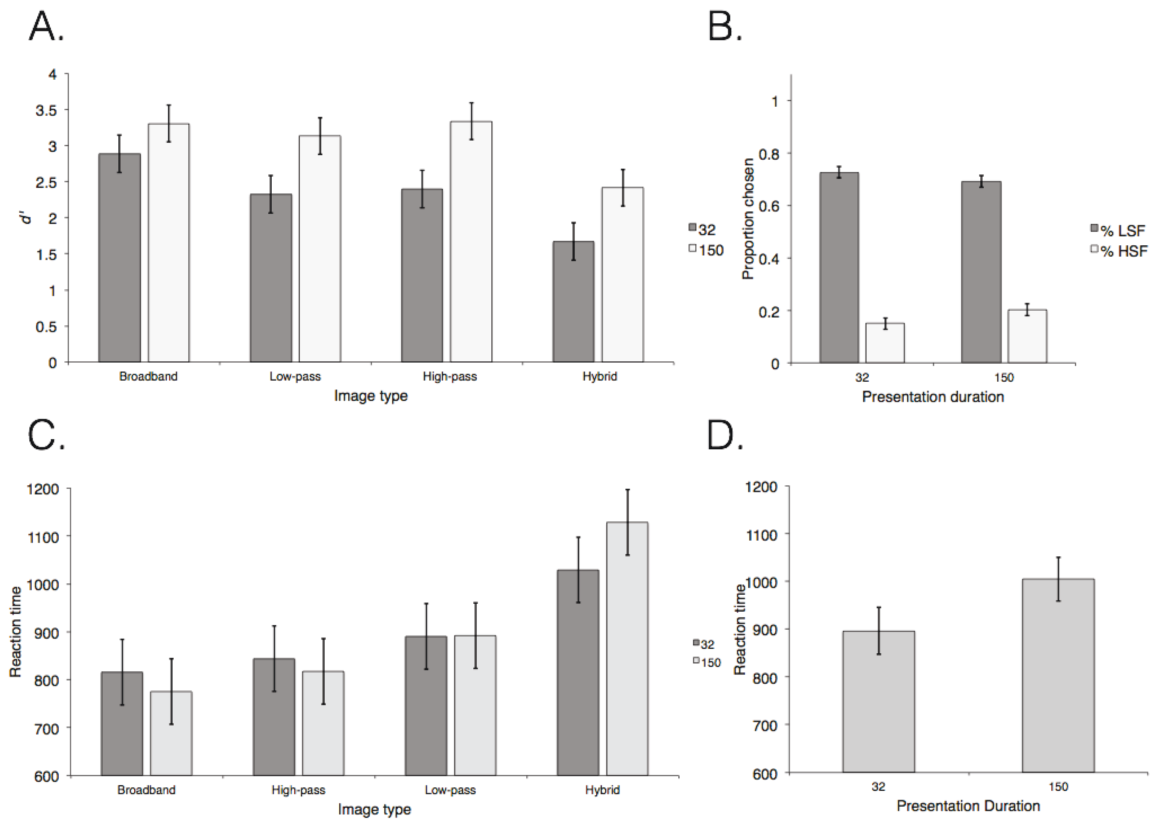


Figure 2.4. The results of Experiment 2. A) Mean *d prime* values for each image type at each presentation duration; B) Percentage of low- and HSF-based hybrid categorization at each presentation duration; C) Mean scene categorization reaction times for each image type at each presentation duration; D) Mean reaction times for LSF-based hybrid categorization at each presentation duration. The error bars represented here are the 95% within subject confidence intervals described by Loftus and Masson (1994).

2.4.3 Discussion

Experiment 2 showed that observers preferred to categorize hybrid images based on LSF content, at both short and long durations. However, an interesting finding is that observers were significantly slower at categorizing hybrid images compared to the other image types. The most likely explanation for this result is that although the probability of a correct answer was greatest for hybrid images (50% versus 25%) their categorization nevertheless led to greater interference effects because they contained competing sources of information. Together with the fact that HSF-based hybrid categorization did not exceed chance performance, and these results corroborate the finding that although observers process information at multiple spatial scales, they nevertheless use a single spatial scale as the basis for categorization (Oliva & Schyns, 1997). Along the same lines, observers in the current study were less sensitive to hybrid images than the other image types. Similar to above, the most parsimonious explanation for this result is that hybrid images differed from control images with respect to the probability of a correct answer. Because observers had a 50% chance at guessing the category of a hybrid image, this essentially reduced the 4-alternative forced choice task to a 2-alternative forced choice task. Thus, although accuracy was comparable between the different image types, sensitivity was nonetheless lower for hybrid images.

The critical finding from Experiment 2 is that observers overwhelmingly preferred to base hybrid image categorization on LSF content, despite the fact that both LSFs and HSFs were perceptually available (Experiment 1). The results of Experiments 1 and 2 thus serve as a baseline for Experiment 3 in which we examined whether we can bias spatial selection by directing attention to either global, or local levels of hierarchical

Navon figures.

2.5 Experiment 3

Experiment 3 was a replication of Experiment 2 with the exception that we asked observers to complete global and local Navon tasks prior to classifying hybrid images. Similar to Experiments 1 and 2, we included control images in order to properly understand how attending locally and globally affected the processing of LSFs and HSFs. Because observers preferred to base hybrid categorization on LSF content, we predicted that LSF-based hybrid categorization would be facilitated following global Navon tasks; that is, LSF-based hybrid categorization would be faster following global Navon tasks than local Navon tasks. Also, because there was no interaction between presentation duration and categorization performance in Experiments 1 and 2, we simplified our design by presenting images at only 32 msec.

2.5.1 Method

2.5.1.1 Observers. Fourteen naïve undergraduate students from Concordia University participated in this study in return for partial course credit. All observers self-reported normal or corrected-to-normal vision.

2.5.1.2 Stimuli and apparatus. Stimuli and apparatus were the same as in Experiment 1 with the following exceptions.

***2.5.1.2.1 Navon task**.* Stimuli used in the Navon task were white Navon letters (RBG values, [255 255 255]; luminance of 102 cd/m$^{2)}$ presented on a gray background (RBG values, [128 128 128]; luminance of 52 cd/m^2). The display consisted of two Navon letters, one in the left and one in the right visual field, located 1° from a centrally located fixation cross. The global and local features of the Navon stimuli were either consistent (e.g., a large C comprised of copies of smaller Cs) or conflicting (e.g., a large T comprised of copies of smaller Cs). The letters used were C, E, H, and T, in all their global and local combinations. Each local letter subtended 0.7˚ X 0.7˚ of visual angle whereas the global letter subtended 5.7˚ X 4˚ of visual angle.

2.5.1.2.2 Scene categorization task. Stimuli in the scene categorization task were the same as in Experiment 1.

2.5.1.3 Procedure.

2.5.1.3.1 Navon task. Trials began with a fixation cross located at the centre of the screen, presented for 250 msec, immediately followed by the presentation of the Navon letters, presented for 100 msec. The task of the participant was to indicate whether the local (local phase) or the global (global phase) configurations of the Navon letters matched. We instructed observers to press the "1" key on the keyboard number pad if they believed that the two Navon letters matched; we instructed observers to press the "2" key if they believed that they did not. Responses were speeded, and no trial-by-trial feedback was provided.

2.5.1.3.2 Scene categorization task. Each trial began with a fixation cross located in the centre of the screen presented for 250 msec, followed by a single natural image presented for a display-to-mask SOA of 32 msec. A mask (the same white noise mask used in the previous experiments) followed image offset and was presented for 64 msec. The image was a broadband, low-pass, high-pass, or a hybrid image. Immediately following offset of the mask, observers were presented with a display screen in which they were asked to indicate the category of the image presented (e.g., city = 1, highway = 2, living room = 3, or valley = 4) by pressing the corresponding number of the category.

The options were presented in the centre of the screen and stayed visible until a response was made. Trial-to-trial feedback was not provided.

2.5.1.4 Design. Observers completed two phases: a local phase and a global phase. In both phases, observers completed both the Navon task and the scene categorization task on each experimental trial (e.g., Navon task – scene categorization task – Navon task – scene categorization task; Martin, Slessor, Allen, Philips, & Darling, 2012). The only difference between the phases was whether observers were asked to indicate whether the local (local phase) or the global (global phase) configurations of the Navon letters matched. An example of a trial type is presented in Figure 2.5. There were an equal number of consistent and inconsistent Navon letters presented. The order in which observers completed the phases was counterbalanced across observers. There was a minimum of 30 minutes and a maximum of 60 minutes between phases. This was done to minimize any potential for interference between the different Navon tasks. Before the start of each phase, observers completed 30 practice trials in order to familiarize themselves with the task. Scene categories used during the practice trials were not used in the experimental trials (e.g., forests and barn scenes) and were not analyzed. Each phase consisted of 16 blocks of 48 trials for a grand total 768 trials.

To ensure that observers were primed to the appropriate attention level from the beginning of both the local and the global phases, observers first completed a respective block (48 trials) of either the global, or local Navon task. Similar to the practice trials, the main purpose of this priming block was to minimize any interference effects from the previous block. Trials in this phase were not analyzed.

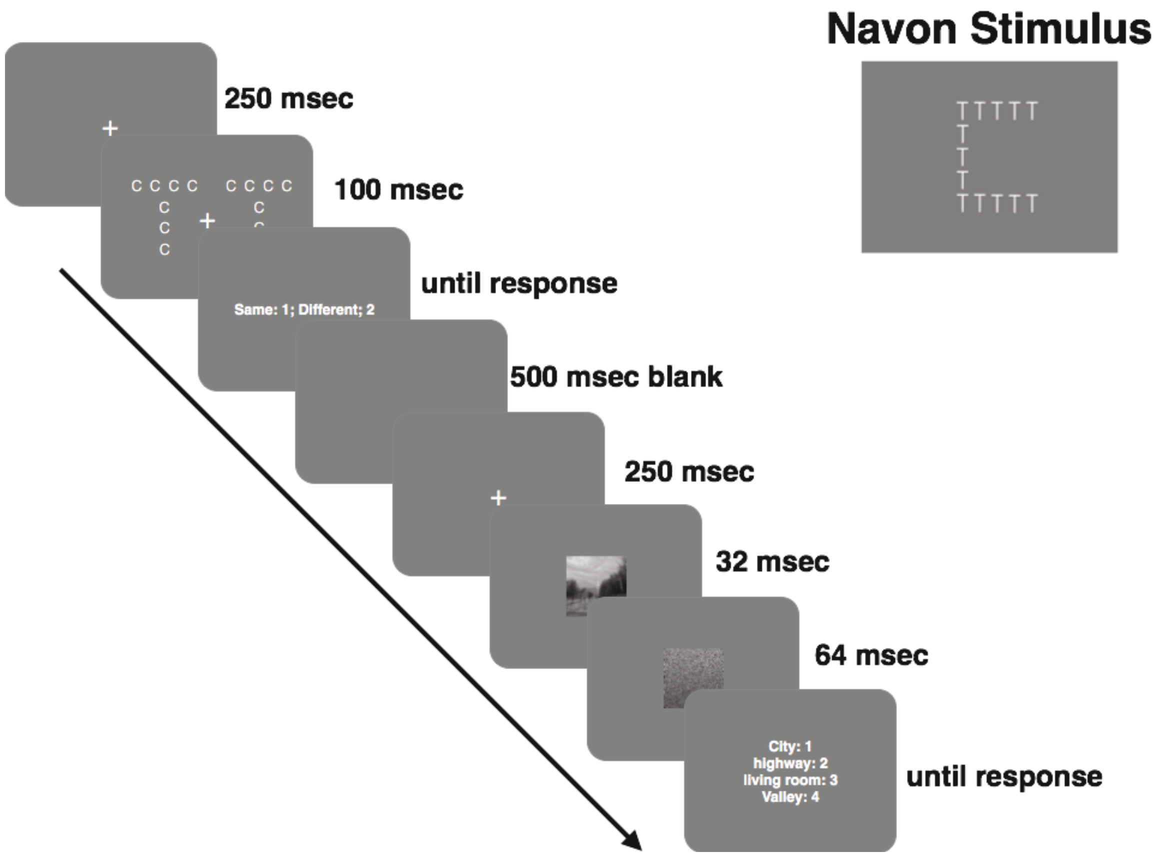


Figure 2.5. An example of a trial sequence in Experiment 3.

2.5.2 Results

2.5.2.1 Scene categorization results. The primary objective of Experiment 3 was to understand how attending to local and global levels of Navon figures affected the subsequent selection of spatial scale information in subsequently presented hybrid images. However, and similar to Experiments 1 and 2, it was necessary that we first understood how attention to hierarchical level affected the processing of low and HSFs within our scenes. Accordingly, we first analyzed sensitivity and RT data between the control images.

2.5.2.1.1 Sensitivity. Mean *d' values* were computed for each trial type. Trial type varied according to image type and Navon processing. These means are displayed in Figure 2.6A and were entered into a 2 (Navon) X 4 (image type) repeated measures ANOVA. There was a significant main effect of image type, $F(3, 39) = 40.15$, $p < .001$, $\eta^2 = .75$, but neither the main effect of Navon nor the image type X Navon interaction were significant, $F(1, 13) = .851$, $p > .373$, $\eta^2 = .06$ and $F(3, 39) = .027$, $p > .994$, $\eta^2 = .02$. Similar to Experiment 2, observers were less sensitive to hybrid images ($M = 1.46$; $SD = 0.27$) than the other image types ($M = 2.44$; $SD = 0.11$) (Ψ_1). Furthermore, observers were more sensitive to broadband images ($M = 2.90$; $SD = 0.42$) than low-pass and high-pass filtered images ($M = 2.21$; $SD = 0.22$) (Ψ_2). There was no difference in sensitivity between low-pass and high-pass filtered images (Ψ_3). As in Experiment 2, the effect sizes associated with Ψ_1 ($\eta^2 = .88$) and Ψ_2 ($\eta^2 = .73$) were similar, replicating the result that observers were most sensitive to broadband images and least sensitive to hybrid images. The effect size between low-pass and high-pass filtered images was relatively smaller ($\eta^2 = .07$). *d'* statistics and contrast analyses are displayed in Table 2.5.

The proportion of low- and HSF-based hybrid categorization is displayed in Figure 2.6B. As in Experiment 2, observers preferred to classify hybrid images based on LSF content in both local and global conditions. Furthermore, HSF-based hybrid categorization was no better than chance in the global condition (M = .22; SD = 0.02), $t(13) = 1.61$, $p > .133$, Cohen's d = .14, and worse than chance in the local condition (M = M = .17; SD = .04), $t(13) = 3.83$, $p < .001$, Cohen's d = .51. Thus, we restricted our analysis to trials on which hybrid categorization was based on LSF content. LSF-based hybrid categorization was higher following global (M = .69; SD = 0.11) than local Navon tasks (M = .62; SD = 0.11), $t(13) = 4.29$, $p < .001$, Cohen's d = 1.14, a difference of .07 (SD = .06; 95% CI [.04, .10]).

Table 2.5. *d prime* statistics for each trial type in local and global conditions in Experiment 3. *d prime* mean difference contrasts in Experiment 3.

	d'							
	Local			Global				
Trial Type	***M***	***SD***	**95% CI**	***M***	***SD***	**95% CI**		
Broadband	2.84	0.52	[2.56, 3.14]	2.95	0.56	[2.63, 3.27]		
Low-pass	2.18	0.37	[1.97, 2.38]	2.25	0.22	[2.12, 2.37]		
High-pass	2.19	0.36	[1.98, 2.39]	2.25	0.21	[2.11, 2.36]		
Hybrid	1.41	0.29	[1.24, 1.58]	1.49	0.28	[1.34, 1.66]		
				Contrasts				
Contrast	***df***	***F***	***p***	M_D	***SE*** **(M_D)**	**95% CI (M_D)**	η^2	
Ψ_1	(1, 13)	94.06	< .001	0.98	0.09	[0.62, 1.36]	.88	
Ψ_2	(1, 13)	17.27	< .001	0.69	0.16	[0.09, 1.28]	.73	
Ψ_3	(1, 13)	1.01	> .336	< 0.01	0.01	[-0.01, 0.03]	.07	

Ψ1: *d'* comparison between hybrid images and the other image types.
Ψ2: *d'* comparison between broadband images and high-pass and low-pass filtered images.
Ψ3: *d'* comparison between low-pass and high-pass filtered images.

2.5.2.1.2 Reaction time. Mean RTs were computed as in the sensitivity analysis and are displayed in Figure 2.6C. We entered these means into a 2 (Navon) X 4 (image type) repeated measures ANOVA. There were significant main effects of image type, $F(3, 39) = 16.15$, $p < .001$, $\eta^2 = .55$, and Navon, $F(1, 13) = 98.55$, $p < .001$, $\eta^2 = .88$. The Navon X image type interaction was not significant, $F(3, 39) = 2.07$, $p > .121$, $\eta^2 = .14$. Reaction times were overall faster following global ($M = 555.41$; $SD = 59.94$) than local Navon tasks ($M = 860.22$; $SD = 94.38$), a difference of 304.81 msec ($SD = 110.70$; 95% CI [211.31, 398.31]).

As in Experiment 2, the significant image type main effect appears to be due to the fact that RTs were slower in response to hybrid images. The contrast comparing RTs between hybrid image types ($M = 835.87$; $SD = 67.77$) and the other image types ($M = 665.14$; $SD = 33.88$) was significant (Ψ_1). Furthermore, RTs were faster for broadband images ($M = 622.15$; $SD = 78.86$) than low and high-pass filtered images ($M = 686.63$; $SD = 34.33$), corroborating the result from Experiment 2 (Ψ_2). There was no significant difference between low-pass ($M = 685.89$; $SD = 50.01$) and high-pass filtered images ($M = 687.37$; $SD = 82.49$) (Ψ_3). As in the previous experiments, the associated effect sizes mirrored the statistical significance results. The largest effect sizes were for Ψ_1 ($\eta^2 = .81$) and Ψ_2 ($\eta^2 = .35$), corroborating the finding that observers were overall fastest to respond to broadband images and slowest to respond to hybrid images. Furthermore, and similar to the sensitivity analysis, the effect size for the comparison between low-pass and high pass filtered images was small ($\eta^2 < .01$), corroborating the finding that there were no meaningful differences between these image types. Reaction time statistics and contrast analyses are displayed in Table 2.6.

As in Experiment 2, we compared RTs between trials on which hybrids were classified according their LSF content (Figure 2.6D). LSF-based hybrid categorization was faster following global compared to local Navon tasks, $t(13) = 6.71$, $p < .001$, Cohen's $d = 1.79$, a difference of 322.32 msec ($SD = 173.12$; 95% CI [222.38, 422.26]).

Table 2.6. Reaction time statistics for each image type in local and global conditions in Experiment 3. Reaction time mean difference contrasts in Experiment 3

	Reaction Time (msec)					
	Local			Global		
Trial Type	***M***	***SD***	**95% CI**	***M***	***SD***	**95% CI**
Broadband	751.40	80.49	[704.93, 797.86]	492.89	110.74	[428.96, 556.83]
Low-pass	830.50	78.04	[785.45, 875.59]	541.29	59.45	[506.96, 575.61]
High-pass	847.93	136.12	[769.34, 926.51]	526.81	72.29	[485.08, 568.54]
Hybrid	1011.06	150.13	[924.39, 1097.74]	660.65	70.20	[620.12, 701.18]
Hybrid - LSF	902.86	86.55	[852.89, 952.82]	580.53	86.55	[530.56, 630.49]

Contrasts

Contrast	***df***	***F***	***p***	M_D	$SE\ (M_D)$	**95% CI** (M_D)	η^2
Ψ_1	(1, 13)	55.61	< .001	170.71	101.65	[112.04, 282.79]	.81
Ψ_2	(1, 13)	7.02	< .021	64.48	23.44	[13.83, 115.13]	.35
Ψ_3	(1, 13)	.002	> .965	1.47	117.89	[-66.58, 69.53]	< .01

Ψ1: RT comparison between hybrid images and the other trial types.
Ψ2: RT comparison between broadband images and high-pass and low-pass filtered images.
Ψ3: RT comparison between high-pass and low-pass filtered images.

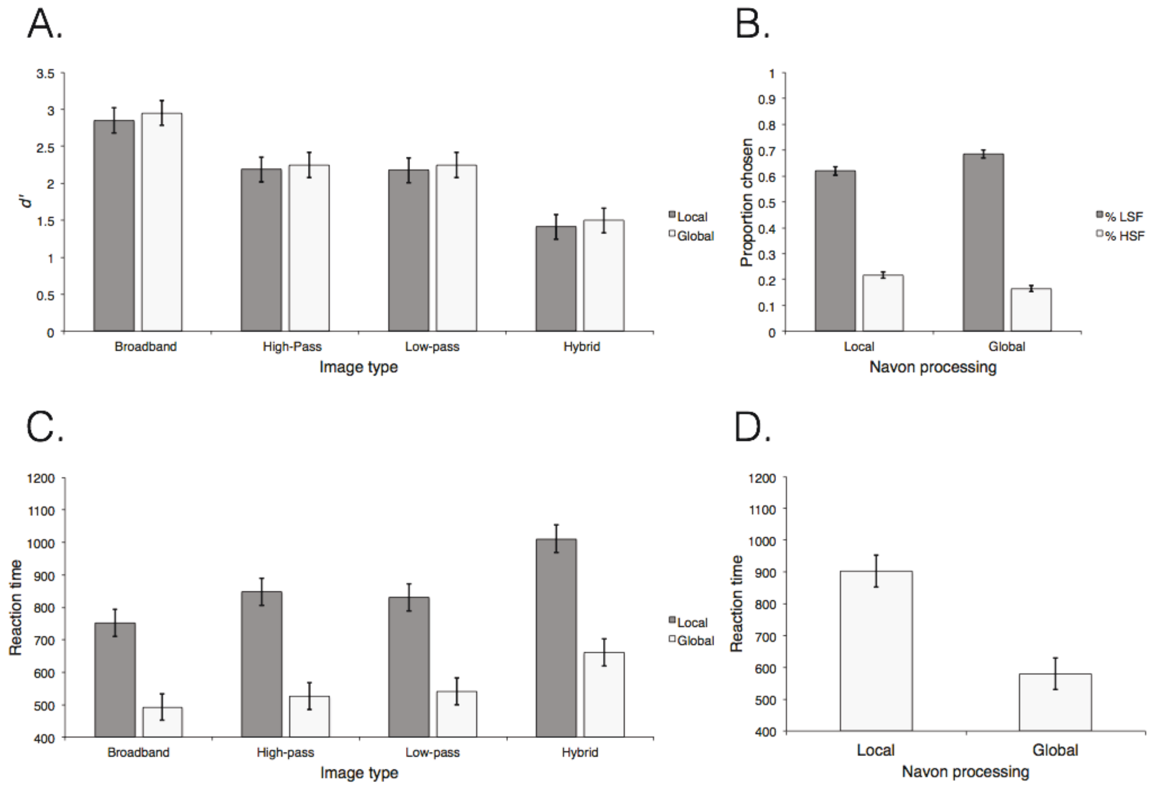


Figure 2.6. The results of Experiment 3. A) Mean *d prime* values for each image type in local and global conditions; B) Percentage of low- and HSF-based hybrid categorization in local and global conditions; C) Mean scene categorization reaction times for each image type in local and global conditions; D) Mean reaction time for LSF-based hybrid categorization following local and global Navon tasks. The error bars represented here are the 95% within subject confidence intervals described by Loftus and Masson (1994).

2.5.2.2 Navon results

2.5.2.2.1 Accuracy. Mean accuracy was computed for the trial types described above (Figure 2.7A). Overall, accuracy was above 90% in all conditions. We entered mean accuracy into a 2 (Navon) X 4 (image type) repeated measures ANOVA. The main effects of image type and Navon were not statistically significant, $F(3, 39) = 2.25$, $p > .098$, $\eta^2 = .15$, and $F(1, 13) = .126$, $p > .728$, $\eta^2 = .01$. Furthermore, the Navon X image type interaction was also not significant, $F(3, 39) = 1.19$, $p > .326$, $\eta^2 = .09$.

2.5.2.2.2 Reaction time. Mean RTs were computed as in the accuracy analysis and are displayed in Figure 2.7B. We entered these means into a 2 (Navon) X 4 (image type) repeated measures ANOVA. There was a significant main effect of image type, $F(3,39) = 6.88$, $p < .001$, $\eta^2 = .35$, and Navon, $F(1, 13) = 56.28$, $p < .001$, $\eta^2 = .81$. The Navon X image type interaction was not significant, $F(3, 39) = .449$, $p > .719$, $\eta^2 = .03$. Overall, global Navon tasks ($M = 379.74$; $SD = 140.10$) were completed faster than local Navon tasks ($M = 567.07$; $SD = 146.40$), This difference was approximately 187.34 msec ($SD = 90.03$; 95% CI [111.29, 263.38]) and corroborated the robust finding of the global precedence effect (Navon, 1977). The main effect of image type appears to be driven by the fact that Navon RTs were overall slowest when completed in conjunction with low-pass filtered images. A significant contrast comparing Navon RTs between low-pass filtered image trials ($M = 505.42$; $SD = 148.53$) and the other image trials ($M = 562.73$, $SD = 132.82$) confirmed this interpretation, $F(1, 13) = 59.14$, $p < .001$, $\eta^2 = .42$. This difference was approximately 57.31 msec ($SD = 25.99$; 95% CI [41.52, 73.09]). The contrast comparing Navon RTs between broadband image trials ($M = 445.04$, $SD = 130.62$) and the combined mean of high-pass filtered image trials and hybrid image trials

($M = 471.58$; $SD = 136.61$) was not significant, corroborating this conclusion, $F(1, 13) = 4.17$, $p > .071$, $\eta^2 = .24$. This difference was approximately 26.54 msec ($SD = 46.85$; 95% CI [-0.05, 53.58]).

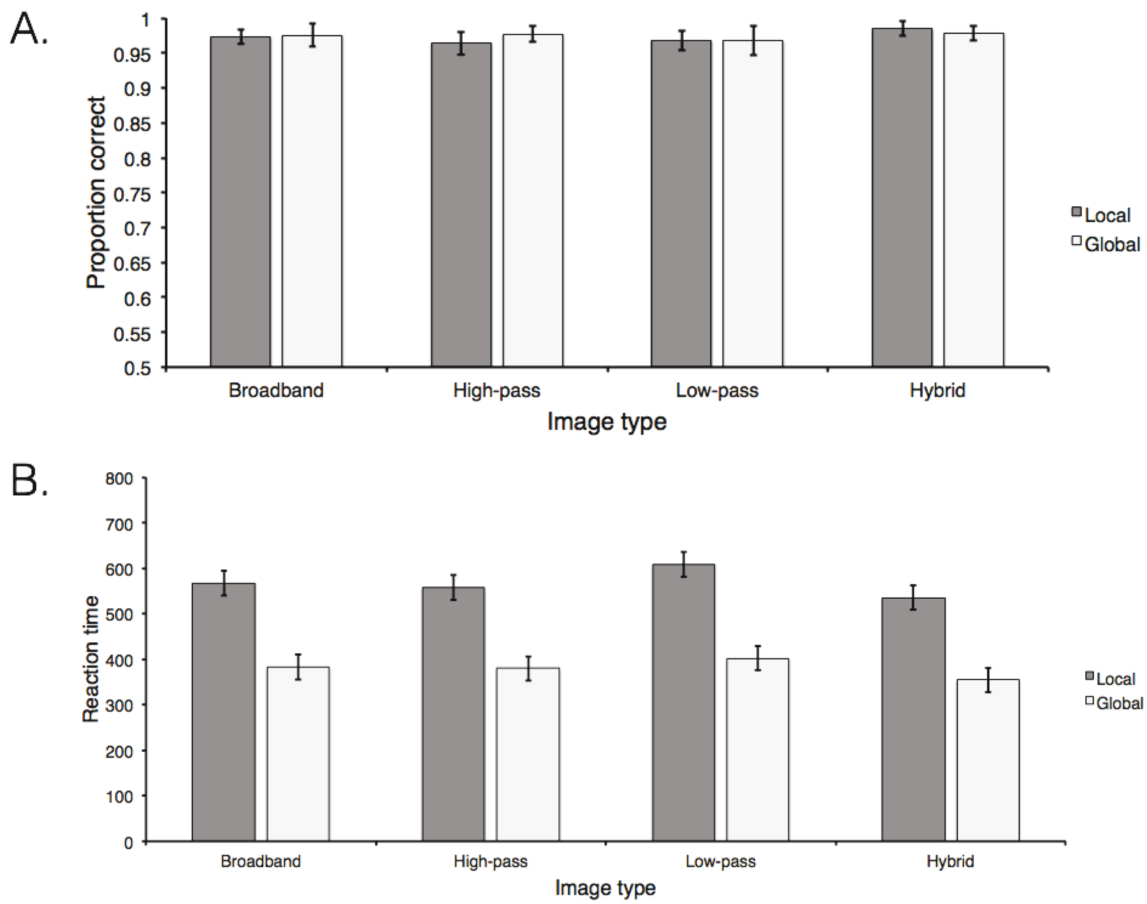


Figure 2.7 Navon results in Experiment 3. A) Mean Navon accuracy for each image type in local and global conditions; B) Mean Navon reaction times for each image type in local and global conditions. The error bars represented here are the 95% within subject confidence intervals described by Loftus and Masson (1994).

2.5.3 Discussion

Global Navon tasks were completed faster than local Navon tasks in Experiment 3, corroborating the global precedence effect (Navon, 1977). A critical result from Experiment 3 is that when given the choice between differing sources of information, observers preferred to categorize hybrids based on LSF content, irrespective of the Navon task completed. Similar to Experiment 2, HSF-based categorization was no better than chance. Consistent with our hypothesis, LSF-based hybrid image categorization was statistically significantly faster following global than local Navon tasks. These results can be interpreted to suggest that we replicated Flevaris and coworkers' (2011) finding that attending globally facilitated the selection LSFs in our hybrid images. However, this interpretation is inconsistent with the finding that both low-pass and high-pass filtered images were categorized faster following global Navon tasks. If attending locally and globally facilitated HSF and LSF processing, respectively, then high-pass filtered images should have been identified faster following local Navon tasks. According to Flevaris and colleagues, however, the selection of spatial frequencies is relative. Thus, although low-pass and high-pass filtered images have HSFs and LSFs removed there are nevertheless still LSFs and HSFs within both image types. Thus, it is possible that the processing of LSFs associated with global Navon tasks facilitated the relatively lower spatial frequencies in both low-pass and high-pass filtered images. This explanation seems likely given that observers preferred to categorize hybrid images based on LSF content. A prediction of this account is that removing a Navon's LSFs should eliminate the benefit associated with categorization following global Navon tasks. In Experiment 4, we directly tested this hypothesis by replicating Experiment 3 with the exception that we used contrast balanced Navon stimuli.

2.6 Experiment 4

Experiment 4 was a replication of Experiment 3 with the exception that Navon stimuli were contrast balanced to suppress LSF information and encourage observers to use HSFs to accomplish both local and global Navon tasks. We confirmed that LSFs were reduced in the stimuli used in Experiment 4, by calculating the log-power spectra and rotationally averaged log amplitude spectra of both the contrast balanced and original Navon stimuli. These analyses are described in Appendix 2.1. We predicted that forcing observers to use HSFs to complete Navon tasks, irrespective of attended level, would eliminate the global advantage associated with scene categorization observed in Experiment 3.

2.6.1 Method

2.6.1.1 Observers. Fifteen naïve undergraduate students from Concordia University participated in this study in return for partial course credit. All observers self-reported normal or corrected-to-normal vision.

2.6.1.2 Stimuli, apparatus, and procedure. Stimuli, apparatus, and procedure were the same as in Experiment 3, expect that Navon stimuli were contrast balanced, such that darker lines surrounded the white lines of the local letters. An example of a trial sequence and a contrast balanced Navon stimulus is displayed in Figure 2.8.

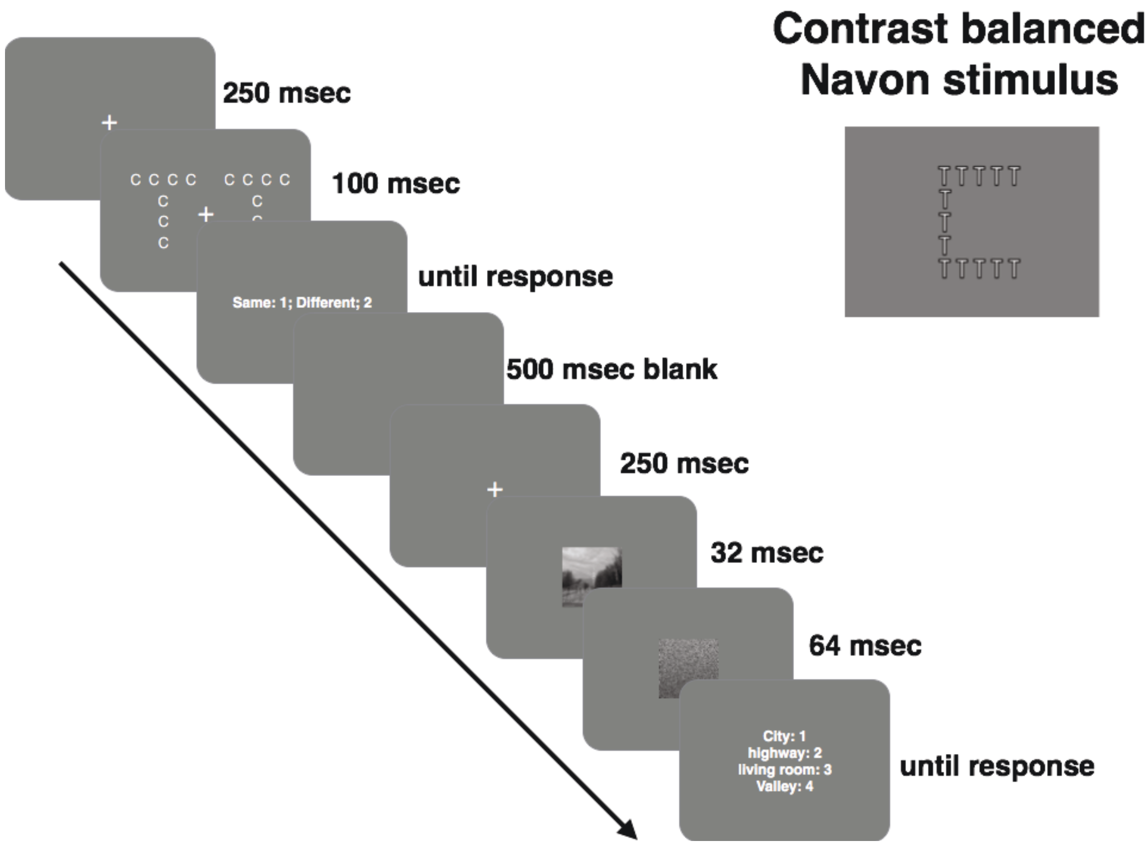


Figure 2.8. An example of a trial sequence in Experiment 4.

2.6.2 Results

2.6.2.1 Scene categorization results

2.6.2.1.1 Sensitivity. *d′* values were computed as in Experiment 3 and are displayed in Figure 2.9A. Overall sensitivity was high, replicating performance in the previous experiments. We entered *d′* means into a 2 (Navon) X 4 (image type) repeated measures ANOVA. There was a significant main effect of image type, $F(3, 42) = 15.41$, $p < .001$, $\eta^2 = .52$, but neither the main effect of Navon nor the image type X Navon interaction was significant, $F(1, 14) = 1,14$, $p > .304$, $\eta^2 = .08$ and $F(3, 42) = .196$, $p > .898$, $\eta^2 = .14$. Similar to previous experiments, observers were less sensitive to hybrid images ($M = 1.75$; $SD = 0.29$) than the other image types ($M = 2.53$; $SD = 0.14$) (Ψ_1). Furthermore, there was no difference in sensitivity between broadband images ($M = 2.64$; $SD = 0.13$) and low-pass and high-pass filtered images ($M = 2.47$; $SD = 0.11$) (Ψ_2). Furthermore, the effect size measures mirrored the statistical significance results. The effect size for Ψ_1 ($\eta^2 = .79$) was larger than the effect size for Ψ_2 ($\eta^2 = .09$), corroborating the finding that observers were less sensitive to hybrid images and equally sensitive to all other image types in Experiment 4. *d′* statistics and contrast analyses are displayed in Table 2.7.

Similar to the previous experiments, observers preferred to categorize hybrid images based on LSF content in both local and global conditions (Figure 2.9B). Furthermore, HSF-based hybrid categorization was worse than chance following both local and global Navon tasks, $t(14) = 12.43$, $p < .001$; and $t(14) = 6.14$, $p < .001$. In contrast to experiment 3, LSF-based hybrid categorization was higher following local than global Navon tasks, $t(13) = 3.93$, $p < .001$, Cohen's $d = 1.07$, a difference of .06 (*SD*

= .06, 95% CI = [.01, .11]). It is interesting to note that the effect size was consistent with the value reported in Experiment 3 (Cohen's d = 1.14), but is in the opposite direction, suggesting a complete reversal of the effect.

Table 2.7. *d prime* statistics for each image type in local and global conditions in Experiment 4. *d prime* mean difference contrasts in Experiment 4.

	d'						
	Local			**Global**			
Trial Type	***M***	***SD***	**95% CI**	***M***	***SD***	**95% CI**	
Broadband	2.77	0.79	[2.41, 3.13]	2.51	0.62	[2.23, 2.79]	
Low-pass	2.59	0.51	[2.36, 2.82]	2.44	0.19	[2.35, 2.53]	
High-pass	2.50	0.44	[2.31, 2.71]	2.34	0.46	[2.13, 2.55]	
Hybrid	1.85	0.49	[1.62, 2.08]	1.66	0.39	[1.49, 1.84]	
			Contrasts				
Contrast	***df***	***F***	***p***	M_D	$SE(M_D)$	**95% CI**(M_D)	η^2
Ψ_1	(1, 14)	50.35	< .001	0.77	0.11	[0.58, 0.95]	.79
Ψ_2	(1, 14)	1.13	> .269	0.17	0.14	[-0.08, 0.42]	.09

Ψ1: *d'* comparison between hybrid images and the other image types.
Ψ2: *d'* comparison between broadband images and high-pass and low pass filtered images.

2.6.2.1.2 Reaction time. Mean reaction time was computed as in Experiment 3 and is displayed in Figure 2.9C. Mean RTs were entered into a 2 (Navon) X 4 (image type) repeated measures ANONA. There were significant main effects of image type and Navon, $F(3, 42) = 23.56$, $p < .001$, $\eta^2 = .63$ and $F(1, 14) = 20.99$, $p < .001$, $\eta^2 = .60$. The Navon X image type interaction was not statistically significant, $F(3,42) = .942$, $p > .429$, $\eta^2 = .06$. In contrast to Experiment 3, RTs were overall faster following local ($M = 543.14$; $SD = 100.37$) than global ($M = 788.98$; $SD = 100.08$) Navon tasks, a difference of 245.84 msec ($SD = 200.74$; 95% CI [104.89, 386.77]). As in the previous experiments, observers were slower to respond to hybrid image types ($M = 796.71$; $SD = 86.81$) than the other image types ($M = 622.52$; $SD = 28.93$) (Ψ_1). Observers were also faster to respond to broadband images ($M = 588.69$; $SD = 56.34$) than low-pass and high-pass filtered images ($M = 639.42$; $SD = 34.67$) (Ψ_2). In contrast to previous experiments, observers were faster to respond to high-pass filtered images ($M = 606.22$; $SD = 49.23$) than low-pass filtered images ($M = 672.63$; $SD = 52.12$) ($\Psi3$).

The largest effect size in Experiment 4 was for Ψ_1 ($\eta^2 = .80$), corroborating previous experiments that observers are slowest to respond to hybrid images. Furthermore, the effect size for Ψ_2 ($\eta^2 = .36$) was similar to the previous experiments, corroborating the finding that observers were fastest to respond to broadband images. However an interesting finding is that the effect size for Ψ_3 ($\eta^2 = .59$) was relatively higher than those reported in previous experiments, suggesting that whereas there was no difference in RTs between low-pass and high-pass filtered images in Experiments 1 - 3, observers took longer to respond to high-pass filtered images than low-pass filtered images in Experiment 4. Reaction time statistics and contrast analyses are displayed in Table 2.8.

As in previous experiments, we compared LSF-based hybrid categorization RTs between local and global conditions (Figure 2.9D). In contrast to Experiment 3, LSF-based hybrid categorization was statistically significantly faster following local than global Navon tasks, $t(14) = 3.21$, $p < .006$, Cohen's d =.91, a difference of 229.62 (SD = 250.69, 95% CI [84.90, 374.34]). Furthermore, the associated effect size was relatively smaller than in Experiment 3 (Cohen's $d = 1.79$), suggesting that although the effect in Experiment 4 reversed direction, its magnitude is smaller.

Table 2.8. Reaction time statistics for each image type in local and global conditions in Experiment 4. Reaction time mean difference contrasts in Experiment 4.

	Reaction time (msec)					
	Local			Global		
Trial Type	***M***	***SD***	**95% CI**	***M***	***SD***	**95% CI**
Broadband	460.32	111.15	[384.66, 535.98]	717.07	110.99	[641.78, 792.36]
Low-pass	554.38	116.30	[475.49, 633.27]	790.88	115.28	[712.68, 869.07]
High-pass	496.61	87.15	[437.50, 555.72]	715.83	121.62	[633.33, 798.33]
Hybrid - LSF	592.35	125.35	[528.09, 656.39]	821.97	115.91	[758.83, 885.13]
Hybrid	661.26	160.27	[552.55, 769.98]	932.15	155.22	[826.85, 1037.43]

Contrasts

Contrast	***df***	***F***	***p***	M_D	$SE(M_D)$	**95% CI** (M_D)	η^2
Ψ_1	(1, 14)	54.99	< .001	174.19	36.60	[95.68, 252.69]	.80
Ψ_2	(1, 14)	7.83	< .014	50.73	21.45	[4.71, 96.74]	.36
Ψ_3	(1, 14)	20.64	< .001	66.41	23.39	[16.23, 116.59]	.59

Ψ_1: Reaction time comparison between hybrid images and the other image types.
Ψ_2: Reaction time comparison between broadband images and low-pass and high-pass filtered images.
Ψ_3: Reaction time comparison between low-pass and high-pass filtered images.

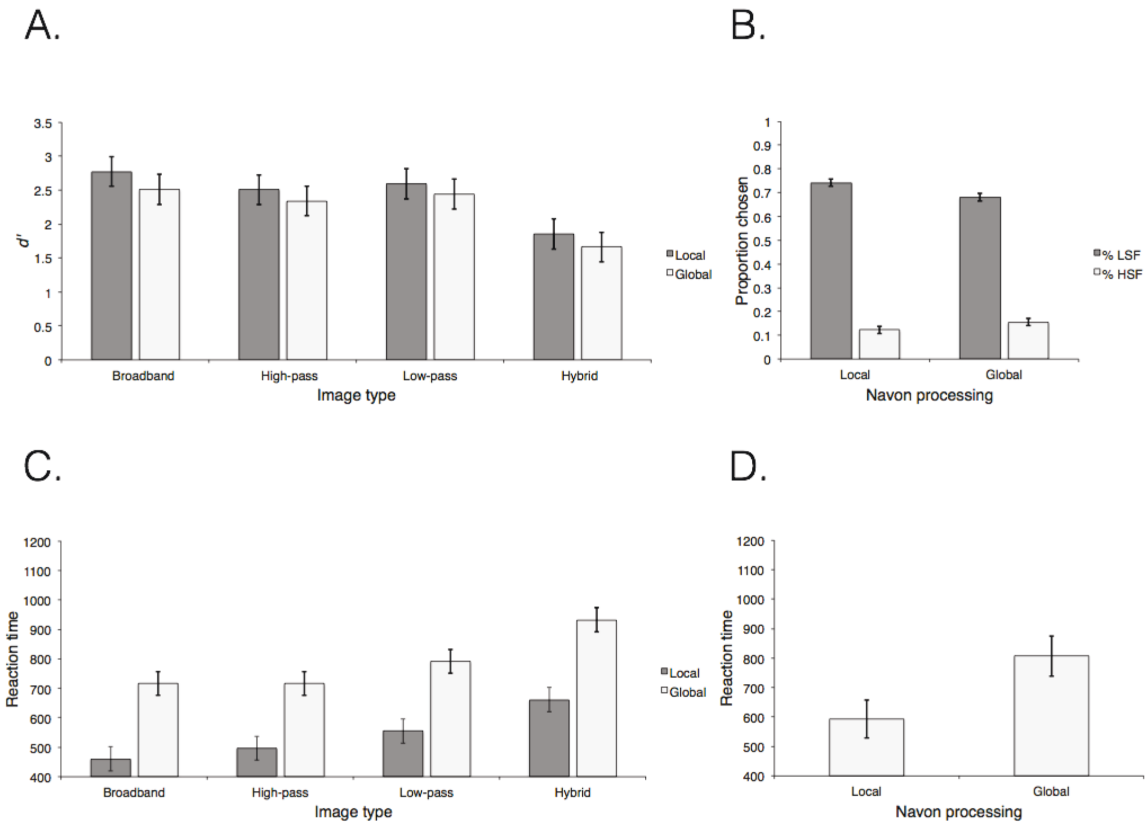


Figure 2.9. The results of Experiment 4. Mean *d prime* values for each image type in local and global conditions; B) Percentage of low- and HSF-based hybrid categorization in local and global conditions; C) Mean scene categorization reaction times in local and global conditions; D) Reaction times for LSF-based hybrid categorization following local and global Navon tasks. The error bars represented here are the 95% within subject confidence intervals described by Loftus and Masson (1994).

2.6.2.2 Navon results

2.6.2.2.1 ***Accuracy.*** Mean accuracy was computed as in Experiment 3 (Figure 2.10A) and replicated the overall high accuracy observed in the previous experiment (> 90%). We compared accuracy by computing a 2 (Navon) X 4 (image type) repeated measures ANOVA. The main effects of Navon and image type were not significant, $F(1, 14) = .736$, $p > .405$, $\eta^2 = .05$ and $F(3,42) = .628$, $p > .601$, $\eta^2 = .04$. The Navon X image type interaction was also not significant, $F(3,42) = .301$, $p > .825$, $\eta^2 = .02$.

2.6.2.2.2 ***Reaction time.*** Mean RTs were computed as in the accuracy analysis and are displayed in Figure 2.10B. We entered group mean RTs into a 2 (Navon) X 4 (image type) repeated measures ANONA. The main effects of Navon and image type were not significant, $F(3,42) = 1.15$, $p > .226$, $\eta^2 = .1$ and $F(1,14) = .924$, $p > .353$, $\eta^2 = .06$. Further, the Navon X image type interaction was not significant, $F(3,42) = 1.52$, $p > .223$, $\eta^2 = .10$.

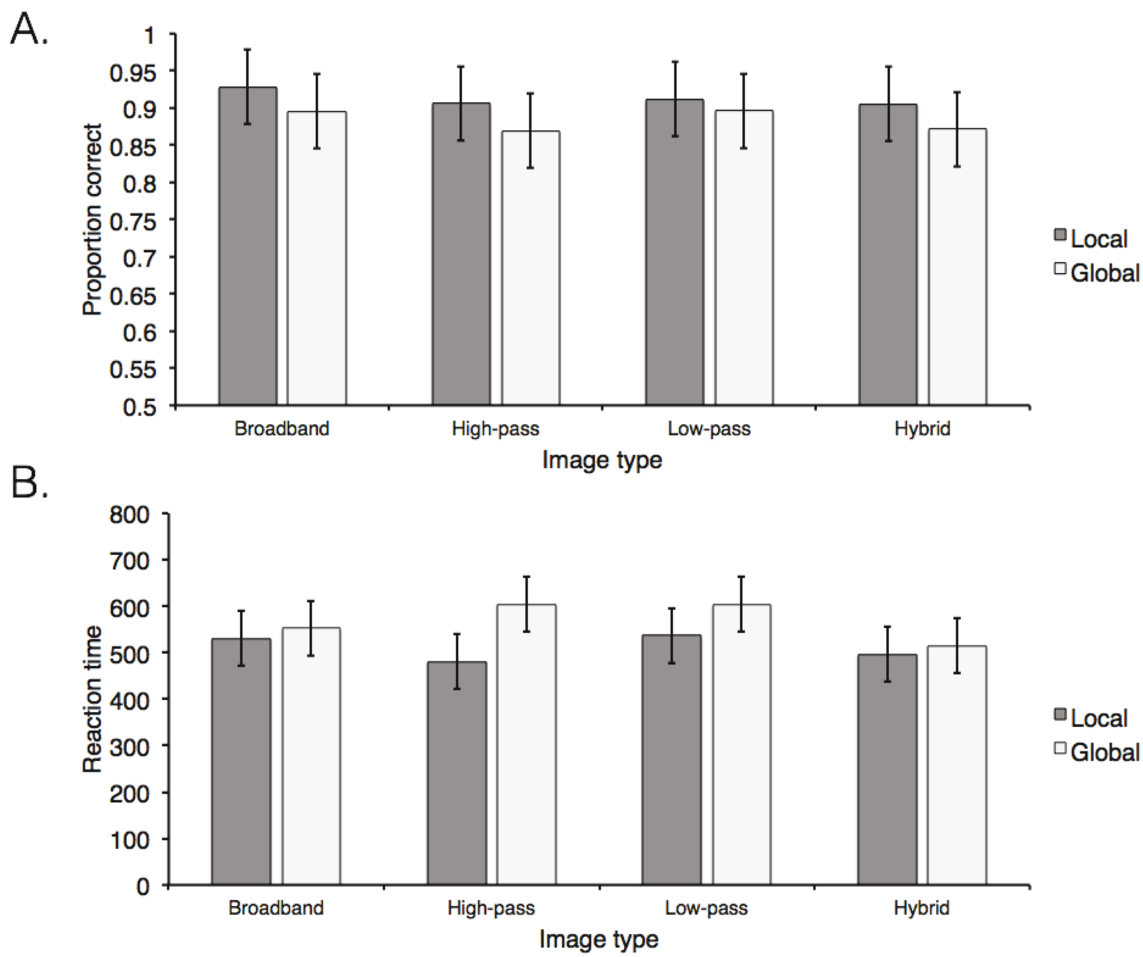


Figure 2.10. Navon results in Experiment 4. A) Mean Navon accuracy for each image type in local and global conditions; B) Mean Navon reaction times for each image type in local and global conditions. The error bars represented here are the 95% within subject confidence intervals described by Loftus and Masson (1994).

2.6.3 Discussion

Experiment 4 was a replication of Experiment 3 with the exception that Navon stimuli were contrast balanced. There was no RT difference between Navon tasks, corroborating the previous finding that contrast balancing Navon stimuli eliminates the global precedence effect (Lamb & Yund, 1993). This forced observers to complete both local and global Navon tasks using HSFs. This afforded the opportunity to determine whether the observed global advantage in Experiment 3 was due to LSF processing associated with global Navon tasks.

An interesting result in Experiment 4 is that observers were faster to respond to high-pass filtered images than low-pass filtered images. One explanation for this result is that suppressing LSFs in Navon stimuli forced observers to complete Navon tasks using HSFs, which in turn, primed the selection of HSFs in high-pass filtered images. As in Experiment 3, observers preferred to categorize hybrid images based on LSF content, following both global and local Navon tasks. However, and in contrast, LSF-based hybrid categorization was slower following global than local Navon tasks. Thus, our prediction that contrast balancing Navon stimuli would eliminate the observed advantage for LSF-based hybrid image categorization following global Navon tasks in Experiment 3 was supported, although we did not predict a complete reversal of the effect. Furthermore, control images were all classified faster following local Navon tasks, suggesting that the global scene categorization advantage in Experiment 3 was due, in part, to the LSFs present in Navon stimuli.

2.7 General Discussion

The four experiments reported in this article investigated how attending to local and global levels of hierarchical Navon figures affected the selection of the spatial scale used for scene categorization. We explored this issue by asking observers to categorize hybrid images immediately following global and local Navon tasks. The composition of hybrid images allows observers to base categorization on either coarse (conveyed by a hybrid's LSFs) or fine (conveyed by a hybrid's HSFs) content. We showed that although observers were sensitive to both types of information (Experiment 1) they overwhelming preferred to base hybrid image categorization on LSF content (Experiments 2 - 4). When hybrid image categorization was not based on LSF content, HSF-based hybrid image categorization was no better (and often worse) than chance. In Experiment 3, we directly examined how attending to global and local levels of hierarchical Navon figures affected LSF-based hybrid categorization, and found that LSF- based hybrid image categorization was faster following global Navon tasks. This corroborates Flevaris and colleagues' (2011) suggestion that attention to the global level of a hierarchical figure facilitates the selection of LSFs. However, inconsistent with Flevaris and colleagues, control images were all categorized faster following global Navon tasks, suggesting that it was not the priming of absolute spatial frequency per se that facilitated LSF-based hybrid image categorization. In Experiment 4, we explored this possibility by replicating Experiment 3 but we forced observers to complete Navon tasks using HSFs, irrespective of the attended level. Similar to Experiment 3, observers preferred to categorize hybrid images based on LSF content. However, and in contrast, LSF-based hybrid image categorization was faster following local Navon tasks, suggesting that LSFs associated with Navon figures were responsible for the scene categorization advantage following global Navon tasks in Experiment 3.

An interesting finding from the present set of studies is that our observers preferred to categorize hybrid images based LSF information in Experiments 2 - 4, despite the fact that they were sensitive to both spatial frequencies in Experiment 1. One possible explanation is that our masking procedure weakened the signal from HSFs more than the signal from LSFs. Such an explanation suggests that our observers preferred to base hybrid image categorization on the spatial frequency with the strongest signal. Consistent with this hypothesis, Losada and Mullen (1995) showed that white noise masks are more effective at masking HSFs than LSFs. Nevertheless, we regard this possibility as unlikely for two main reasons. First, our observers were more sensitive to a hybrid image's HSFs than LSFs in Experiment 1; and second, as mentioned in the discussion of Experiment 1, our masking effects were particularly weak, suggesting that neither the HSF signal nor the LSF signal were strongly affected by our masking procedure. Our preferred interpretation of these apparent conflicting results is that they corroborate previous research that has shown a critical role for LSFs in rapid scene categorization (Loschky & Simons, 2004; McCotter, Gosselin, Sowden, & Schyns, 2005; Schyns & Oliva, 1994; Oliva & Schyns, 1997). The present results provide further evidence for this hypothesis by demonstrating a preference to use LSF information, despite the fact that HSF information is more salient.

A comparison between the present work and the apparent automaticity of scene perception under dual task conditions is particularly relevant. Cohen and colleagues (2011) suggested that attention task difficulty is the reason some studies have documented impaired scene perception (Walker, Stafford, & Davis, 2008) whereas others have not (Li, VanRullen, Koch, & Perona, 2002; Rousselet, Fabre-Thorpe, & Thorpe, 2002). The present work suggests an alternative explanation. Specifically, that impaired scene perception under dual task conditions could be a function of the type of attentional distribution needed to complete the attention task. For example, it seems more likely that a cost of dividing attention would emerge in situations in which the tasks are similar, because the potential for interference from completing the two tasks should be greater. Given that scene categorization was facilitated following global Navon tasks in the present study (at least with unaltered stimuli), suggests the completion of simultaneous attention tasks that require global attention would be more likely to interfere with scene categorization than those that require local attention. Brand, Johnson, and Von Grünau (2012) provided support for this hypothesis by demonstrating that the completion of a concurrent task that requires global attention interferes with scene categorization, but a concurrent task that requires local attention does not.

One issue the present study was unable to resolve is why scene categorization was faster following local Navon tasks in Experiment 4. This is particularly true for hybrid images, as it is unclear how attending locally would facilitate categorization based on LSF content. If LSFs associated with global Navon tasks facilitated LSF-based hybrid categorization in Experiment 3, then removing that information should have eliminated the global benefit, but should not have resulted in a benefit following local Navon tasks. The fact that it did suggests that observers were using different types of information within a hybrid image’s LSF content as the basis for categorization in Experiments 3 and 4, respectively. This conclusion is consistent with Oliva and Schyns' (1997) suggestion that coarse-to-fine information is orthogonal to global-to-local information; that is, there is both coarse and fine information at each spatial scale, and it is possible to direct attention to either level. Consider, for example, the low-pass filtered Navon stimulus in Figure 2.11. The small “c” represents the image’s local features, and the large “T” represents the image’s global feature. According to the global-to-local hypothesis, the fine information in the image (i.e., the small c's) should be unrecognizable because the HSFs that convey that information have been removed. Nevertheless, it is evident in the figure that even though HSFs have been removed, that local information remains. Thus, although observers preferred to categorize hybrids based on LSF information in both Experiments 3 and 4, the selection of a Navon’s LSFs (Experiment 3) and HSFs (Experiment 4) facilitated the selection of different information within a hybrid image’s LSF content. Alternatively, this result can be attributed to a switch cost between the Navon task and the scene categorization task. It is possible that slower responses following local processing in Experiment 4 is due to the fact that the switch from HSF information to LSF information is very time consuming. Conversely, there was no switching in Experiment 3 because observers relied on LSF information for both global and local Navon tasks. Unfortunately, the present study was not designed to identify these differing sources of information.

Figure 2.11. An example of a low-pass filtered Navon stimulus.

Another interesting question arising from the present results is whether a hybrid image's HSFs were encoded in Experiments 2 - 4. Although observers preferred to categorize hybrids based on LSFs in Experiments 2 - 4, the results of Experiment 1 suggest that both spatial scales were perceptually available. This suggestion is consistent with Oliva & Schyns (1997, Experiments 3 and 4) who showed that when a hybrid image's LSF content is the preferred spatial scale, observers nevertheless still process a hybrid image's HSF information implicitly. Along the same lines, de Gardelle and Kouider (2009) found that the non-preferred spatial scale information could facilitate scene categorization. de Gardelle and Kouider asked observers to determine whether a full broadband face presented below conscious awareness was of a famous person. A hybrid face preceded the target face and it was constructed such that either its LSFs, or HSFs corresponded with target identify. The critical point here is that face identification is typically based on the relatively higher spatial frequencies of a hybrid face. Thus, only HSF-hybrid image primes should have facilitated target identification. In contrast, de Gardelle and Kouider reported that both LSF- and HSF-hybrid image primes facilitated target identification. What's more, whereas the effect HSF-hybrid image primes increased significantly with exposure duration, the effect of LSF-hybrid image primes did not. Thus, although LSF information was not preferred, it nevertheless played a small role in categorization, most likely restricted to unconscious processing.

The question relating to the role of attention in scene categorization is currently a major source of debate in psychology. Traditionally, this question is addressed by examining the automaticity of scene perception, and whether or not conscious scene perception can occur in the absence of attention. The present article addressed this question from a different angle. It examined how attention facilitates the selection of information used in scene categorization. Along the same lines, Larson, Freeman, Ringer, and Loschky (2014) showed that manipulations of spatial attention influence the selection of scene information. Similar to the global processing bias in the present study, Larson and colleagues reported that scene categorization is initially based on information originating from central vision, with contributions from peripheral vision emerging later on (i.e., a central-to-peripheral processing bias). Larson and colleagues reported that this central processing bias is reduced when the spatial distribution of attention is manipulated so that it emphasizes information in the periphery. Thus, although Larson et al. did not investigate the interaction between attention and spatial scale processing, their results nevertheless converge with the present results to suggest that one role of attention in scene categorization is to select scene information.

The primary purpose of the present experiments was to address how attention to local and global levels of Navon figures affects the selection of spatial scale information used in scene categorization. This investigation was largely based on the connection between the Navon task spatial scale and the spatial scale used for scene categorization. As such, it is reasonable to assume that the categorization of different scene types could also differentially affect the completion of the Navon task. The results of Experiments 3 and 4 allude to this possibility. Whereas Navon processing was slowest when completed in conjunction with low-pass filtered images in Experiment 3, there was no difference in Navon task RTs as a function of scene type in Experiment 4. Although we can only speculate as to the reason for this difference, it appears to be related to the amount of LSFs in the Navon stimuli. Navon stimuli in Experiment 4 were contrast balanced, such that their LSF content was suppressed compared to the Navon stimuli used in Experiment 3. Combined with the fact that LSFs were the preferred spatial scale in all experiments, this suggests that the observed Navon slowing in Experiment 3 following low-pass scene categorization was due, in part, to an increased use of LSFs in Experiment 3 compared to Experiment 4.

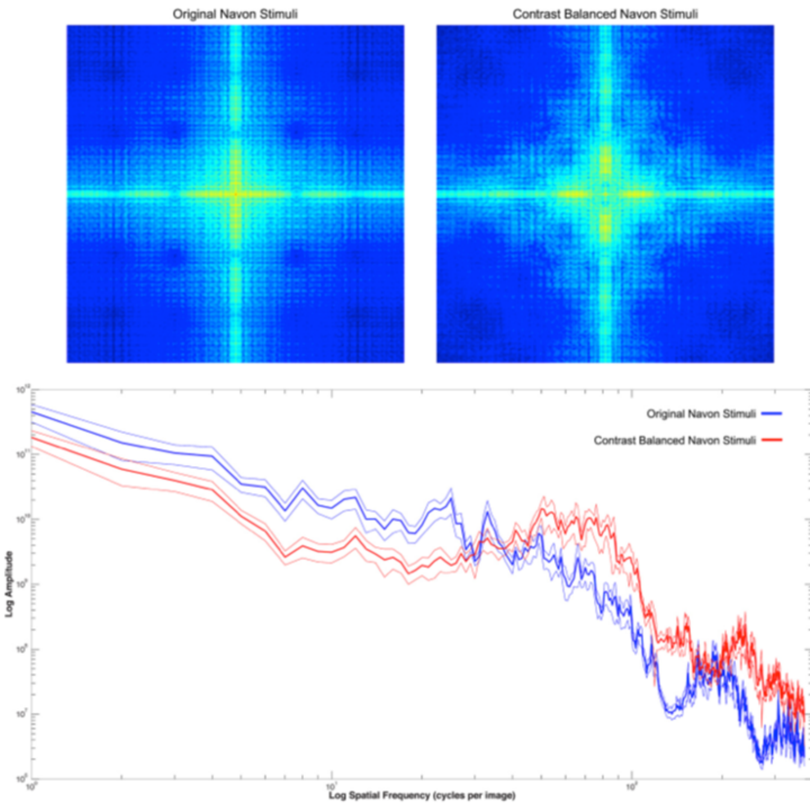
In conclusion, the present set of experiments demonstrates that attending locally and globally affects the selection of spatial scale information used for rapid scene categorization. The present results also converge with previous research in suggesting that LSF information is important in rapid scene categorization (Loschky & Simons, 2004; McCotter, Gosselin, Sowden, & Schyns, 2005; Schyns & Oliva, 1994; Oliva & Schyns, 1997) and extends these findings by demonstrating that the selection of LSF information is affected by manipulations of attention. Thus, although the present results do not conclusively demonstrate that scene perception requires attention, they nevertheless suggest that attention plays a role in facilitating the selection of scene information.

Appendix 2.1 Chapter 2: Reduction of Low Spatial Frequency Content in Contrast Balanced Navon Stimuli.

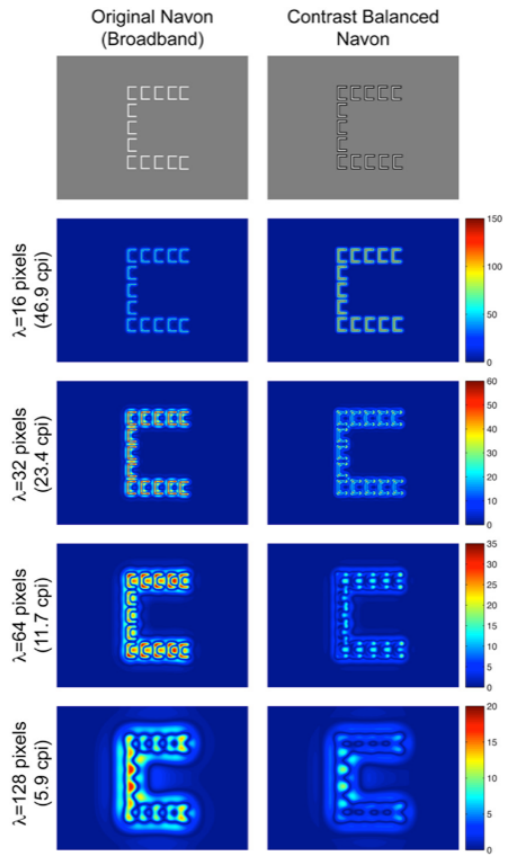
Previous researchers used contrast balanced Navon stimuli to suppress (Lamb & Yund, 1993) the low spatial frequencies contained within Navon stimuli. Implementing contrast balanced Navon stimuli encourages observers to use only the remaining high spatial frequencies to accomplish both the local and global Navon letter tasks. To verify that the low spatial frequency content was reduced in the stimuli used for Experiment 4, we calculated the log-power spectra and rotationally averaged log amplitude spectra of both the contrast balanced and original Navon stimuli.

As can be seen in supplementary Figure 2.1, the balancing of contrast across the edges of local elements of the Navon stimuli has the effect of reducing the overall amplitude at the low spatial frequencies, while increasing the amplitude at the high spatial frequencies. Therefore, the addition of the borders to the contrast balanced Navon stimuli is not causing a masking effect of the higher spatial frequencies on the lower spatial frequencies, as there is a physical reduction in the low spatial frequency content.

One possible explanation for this reduction in low spatial frequency content is due to the Fourier analysis introducing an artifact into the stimuli. To exclude this possibility, we convolved the stimuli with a bank of log Gabor filters in the spatial domain. Log Gabor filters were created in Mathworks Matlab (ver. 2013b), using a starting minimum wavelength of the filter to be 16 pixels (or 46.9 cpi). Each filter was rendered at 6 possible orientations (0-150° in 30° increments), with the final response at each spatial frequency being created by averaging across all orientations. Each subsequently lower spatial frequency filter doubled the wavelength, creating a total of four spatial frequencies (46.9, 23.4, 11.7, 5.9 cpi). As can be seen in supplementary Figure 2.2, at the highest spatial frequency (46.9 cpi), the contrast balanced Navon stimuli show a stronger response relative to the original Navon stimuli. However at lower spatial frequencies (23.4, 11.7, and 5.9 cpi), the contrast balanced Navon stimuli show a weaker response in comparison to the original Navon stimuli. We therefore conclude that the reduction of the low spatial frequency component introduced by the contrast balanced Navon is not an artifact, but instead represents a quantifiable reduction in low spatial frequency content.



Supplementary Figure 2.1. Top) Log-power spectra for the original and contrast balanced Navon stimuli (averaged over 16 stimuli used in Experiments 3 & 4). In Fourier space, low spatial frequencies are located toward the center of the image, with increasing spatial frequency content towards the image edge. Bottom) Log amplitude spectra for stimuli, averaged across orientation, with 95% confidence intervals.



Supplementary Figure 2.2: Example of the original and contrast-balanced Navon stimuli convolved with a bank of log Gabor stimuli of different spatial frequency wavelength (l, in pixels), with corresponding cycles per image (cpi). Colour bars represent response of the filter at each spatial frequency, with red depicting a strong response.

Chapter 3: The effects of distributed and focused attention on rapid scene categorization

John Brand & Aaron P. Johnson

3.1 Abstract

It is argued that distributed attention facilitates the rapid extraction of summary statistics, which in turn, underpins rapid scene categorization (Evans & Treisman, 2005). In the present set of studies, we directly examined this hypothesis by investigating whether distributed, or focused attention is more compatible with the extraction of both summary statistics (Experiment 1) and semantic scene information (Experiments 2 – 4). Experiment 1 replicated Chong and Treisman's (2005) result that mean circle size judgments are more compatible with a distributed attention task than a focused attention task. Experiment 2 investigated whether this finding extends to simple scene categorization by replacing the averaging task with an animal detection task. Consistent with Experiment 1, the ability to detect the presence of an animal was more compatible with a distributed attention task than a focused attention task. Experiments 3 and 4 addressed whether distributed attention influences more complex scene categorization tasks in the same fashion. When observers were asked to classify scenes based on their basic level (e.g., beach or forest; Experiment 3), there was no statistically significant difference between focused and distributed attention task conditions; however, superordinate level categorization (e.g., natural or manmade; Experiment 4) was faster when combined with a task requiring distributed attention compared to a task requiring focused attention.

3.2 Introduction

Scene perception captures the interest of researchers because of the paradox between the ease of everyday vision, and the severe attentional limitations observed in laboratory studies (Braun, 2003). Although converging evidence from inattentional blindness (Simons & Chabris, 1999), change blindness (Simons & Levin, 1997), and attentional blink (Raymond, Shapiro, & Arnell, 1992) studies demonstrate that significant changes go unnoticed without visual attention, the processing of everyday scenes operates uninterrupted (e.g., the right half of a room does not disappear if you focus on the left; Braun, 2003; Block, 1995; Wolfe, 1999). How is this apparent incongruity reconciled? One commonly cited solution is that conscious scene perception does not require attentional resources (Li, VanRullen, Koch, & Perona, 2002; Rousselet, Fabre-Thorpe, & Thorpe, 2002).

For a process to be considered automatic it is argued that it must be completed rapidly, without intention, and be immune to interference caused by concurrent processing (Brown, Gore, & Carr, 2002). Scene categorization occurs in as little as 120 msec (Thorpe, Fize, & Marlot, 1996; Fabre-Thorpe, Delorme, Marlot, & Thorpe, 2001) and is suggested to occur below the level of conscious awareness (Koch & Tsuchiya, 2007; Tonomi & Koch, 2008). Together, these two findings suggest that scene perception satisfies the first two automaticity requirements. Yet a discrepancy between two sets of research findings questions whether scene perception is immune to congruent processing: evidence for impaired scene perception under dual task conditions on the one hand (Walker, Stafford, & Davis, 2008; Cohen, Alvarez, & Nakayama, 2011), and no evidence of scene categorization impairment when attention resources are allocated to a simultaneous attention task on the other (Li, VanRullen, Koch, & Perona, 2002; Rousselet, Fabre-Thorpe, & Thorpe, 2002).

Li and coworkers (2002) were among the first to demonstrate scene perception in the absence of attention. They asked subjects to compete a dual task in which an animal detection task (e.g., detect the presence of an animal) was presented in the periphery, 53 msec after onset of a centrally presented letter discrimination task (e.g., search for an "L" among rotated "T"s). Li and coworkers instructed their observers to answer both tasks as quickly and as accurately as possible, but to respond first to the scene categorization task, followed by the central task. In a control condition, observers were only asked to respond to the scene categorization task. Li and coworkers reasoned that if scene categorization occurred without the need for attentional resources, then performance should not differ between single (i.e., when attention is available to be allocated to the scene task), and dual task (i.e., when attention is spilt between the two tasks) conditions. Although scene categorization performance was better when completed alone, this difference was not statistically significant, suggesting that scene categorization occurs in the near absence of attention.

According to Evans and Treisman (2005; see also Treisman, 2006; Evans & Chong, 2011), animal detection in the absence of attention is evidence that scene categorization is accomplished by the parallel detection of unbound features (e.g., feathers, wings, and beaks) that define a target category (e.g., a bird). The theory is that basing categorization on a number of unbound features avoids the need for attention to bind them together in order to perceive an experienced whole (see e.g., Feature Integration of Theory of Attention; Treisman & Gelade, 1980). An assumption of this theory is that although observers are able to categorize a briefly presented image (e.g., yes it was an animal), they do not have a full representation of the scene (e.g., it was a bird in the upper left hand corner). Evans and Treisman directly tested this hypothesis using a rapid serial visual presentation (RSVP) paradigm. They asked participants to detect both the presence and location of an animal when either presented alone, or in the presence of a human distractor. Although observers were accurate at identifying the presence of an animal when presented alone, they were nevertheless poor at identifying its location. Furthermore, animal detection was impaired in the presence of a human distractor, suggesting that the presence of shared features among humans and animals reduced animal detection sensitivity.

Consistent with Evans and Treisman (2005), research also suggests that the extraction of shared scene features interferes with target detection. Li, Iyer, Koch, and Perona (2007), for example, reported that observers require longer presentation durations to report the identity of an animal species, compared to when they are asked to indicate the presence of an animal. When viewed within Evans and Treisman's framework, this suggests that whereas the presence of an animal requires the identification of a single feature, the identification of a specific species requires the integration of several features (e.g., a beak identifies a bird). Similarly, Thorpe and Fabre-Thorpe (2002) found that the ability to detect the presence of an animal decreased as the number of scenes presented increased. When observers were asked to determine the presence of an animal in a display consisting of four simultaneously presented scenes, performance was significantly worsened compared to when three or fewer scenes were presented. Finally, Walker, Stafford, and Davis (2008) demonstrated impaired scene categorization under dual task conditions. In one of their studies (Experiment 2a), they briefly presented (170 msec) observers natural images that were superimposed with four centrally located letters, arranged in a square. On single task trials, they asked observers to indicate if the image contained an animal. On dual task trials, they asked observers to first indicate if the four letters contained a vowel before completing the animal detection task. Critically, and in contrast to Li and colleagues (2002), the images used in this study contained one to four objects. The authors found that scene categorization performance was worse on dual task trials than on single task trials when scenes contained four different objects. However, this effect, although still present, was greatly reduced for trials on which the image contained a single object.

In order to explain how disjunctive features contribute to rapid scene categorization, Evans and Treisman (2005) proposed that there are different types of attention, and that each type facilitates the selection of a different type of information (see also, Chong & Treisman, 2005). For example, focused attention is required in order to perceive whole objects, whereas distributed attention is responsible for the extraction of whole set statistical descriptors that underpins the formation of disjunctive features. The theory is that when attention is distributed over a set of similar items (e.g., a set of beach pebbles), the visual system automatically extracts statistical properties of the set (e.g., average size and texture) that can be used to make rapid decisions (e.g., this is a beach scene). This automatic averaging mechanism appears to be general in its operation applying to both low-level features and high-level properties. Within 200 msec, observers can extract the average size of a set of circles (Ariely, 2001; Chong & Treisman, 2003; 2005), average direction of a set of randomly moving dots (Watamaniuk, Sekuler, & Williams, 1989), and average emotion or identity in a set of similar faces (de Fockert & Wolfenstein, 2009; Haberman, & Whitney, 2007, 2009; Haberman, Harp, & Whitney, 2009). Extracting statistical values (e.g., the mean, range and variance) on a number of dimensions facilitates the formation of disjunctive features that can be used in rapid decision-making; thereby allowing for an economical description of a scene, without the need for focused attention (for a review, see Treisman, 2006; Evans & Chong, 2011).

Chong and Treisman (2005) provided support for this hypothesis by demonstrating that the extraction of average circle size was more easily achieved when combined with a distributed attention task compared to a focused attention task. They asked observers to indicate which of two test circles corresponded to the average size of a preceding set of circles. This mean discrimination task was completed in conjunction with an attention task that required either focused (e.g., indicate the orientation of a small rectangular frame located in the centre of the set) or distributed (e.g., indicate the orientation of a large rectangular frame encompassing the display) attention. The dependent variable was the diameter difference needed between the two test circles to achieve 75% correct performance. The mean diameter difference was smaller when observers completed the distributed attention task compared to the focused attention task, suggesting that the computation of mean circle size was more compatible with tasks requiring distributed rather than focused attention.

If distributed attention facilitates the extraction of statistical properties that can be used as the basis for rapid categorization, then it should also facilitate scene categorization directly. In the present set of studies, we explored this hypothesis by using

Chong and Treisman's (2005) paradigm in conjunction with scene categorization tasks. In Experiment 1, we replicated Chong and Treisman's original task. However, and in contrast to their original design, we concentrated on reaction time (RT) measures - rather than threshold differences - so that we could easily compare the results with scene categorization performance. Observers were asked to complete a mean discrimination task while determining the orientation of either a large (distributed attention), or small (focused attention) rectangle. Mean discrimination RTs were faster when observers completed a distributed attention task than a focused attention task, replicating Chong and Treisman's finding. Experiment 2 was a replication of Experiment 1 with the exception that we replaced the mean discrimination task with an animal detection task, which previous studies had used to measure scene categorization ability (see, e.g., Li et al., 2002; Rousselet et al., 2002; Thorpe et al., 1996). Similar to Experiment 1, RTs in response to the presence of animal were faster when combined with a distributed attention task compared to a focused attention task. In Experiments 3 and 4, we investigated whether this effect extends to more complex scenes by replacing the animal detection task with natural scene categorization tasks. In Experiment 3, observers classified natural images based on their basic level (e.g., beach) and in Experiment 4 observers classified natural images based on their superordinate level (e.g., natural). Whereas there was no difference between the effects of distributed and focused attention tasks on basic level categorization, superordinate level categorization was faster when combined with a distributed attention task compared to a focused attention task.

3.3 Experiment 1

The purpose of Experiment 1 was to replicate Chong and Treisman's (2005) finding that judgments about mean circle size are more compatible with a distributed attention task compared to a focused attention task. In contrast to their original investigation, we modified our design to concentrate on RT measures so that we could obtain a baseline pattern of results to which the results of Experiments 2 – 4 could be compared.

3.3.1 Method

3.3.1.1 Observers. Observers were 15 Concordia undergraduate students who received partial course credit for their participation, or $10 in monetary compensation. All observers self-reported normal or corrected-to-normal vision. Data from a single observer was discarded because they failed to follow task instructions. The University Human Research Ethics Committee at Concordia University approved all experiments reported in this article and participants provided informed consent.

3.3.1.2 Stimuli and apparatus. Stimuli were presented on a 21-in. Viewsonic 225fb CRT monitor (1024 X 768 resolution; 100 Hz refresh rate) controlled by a Dell Precision T3400 core2 quad processor running Microsoft Windows 7. Experiment Builder (ver. 1.10.1025; SR Research, Ottawa, Ontario) was used to display the stimuli and record the responses. The stimuli used in the present study were the same stimuli used in Brand, Oriet, and Sykes-Tottenham (2012). They were green circles presented on a white background. The diameters of the circles ranged from 4 to 96 pixels in 4-pixel increments. When viewed from a distance of 54 cm, the minimum and maximum sizes of the circles subtended 0.2° and 4.1° of visual angle, respectively. Critically, Teghtsoonian

(1965) reported that a power function with the exponent of .76 best described the relationship between the actual and perceived size of circles. Thus, sets were carefully constructed to ensure that all circles differed from one another by at least one step on this power function.

3.3.1.3 Procedure. Each trial began with a fixation cross located at the centre of the screen presented for 500 msec, immediately followed by the presentation of 12 uniquely-sized circles, presented for 200 msec. On each trial a small rectangle was always present in the middle of the display and a large rectangle bordered the outside. Similar to Chong and Treisman's (2005) study, the sizes of the rectangles were constructed in order to ensure that the aspect ratios of each rectangle orientation (vertical and horizontal) were the same between the small and large rectangle. The size of the large rectangle was 442 X 642 pixels. The size of the small rectangle was determined by dividing the dimensions of the large rectangle by 6. Thus, the small rectangle measured 74 X 107 pixels. When viewed from a distance of 54 cm, the large rectangle was either 26.0° X 18.0° or 18.0° X 26.0° and the size of the small rectangle was either 3.02° X 4.43 ° or 4.43° X 3.02°.

Immediately following offset of the display, we presented observers with two test circles, which were presented until the participant made a response. The circles appeared directly to the right and to the left of the fixation cross, with the edge of the circles approximately 1° away from fixation. The diameter of one of the test circles always corresponded to the mean diameter size of the preceding set of circles, whereas the other was either smaller or larger. Chong and Treisman (2005) showed that in order for observers to achieve 75% accuracy on the mean discrimination task, they needed at least a 31% diameter difference between the test probes when discriminating the orientation of the large box, and 37% diameter difference when discriminating the orientation of the small box. Thus, the foil in our study was always at least 37% larger, or smaller than the mean size test probe. We asked observers to as quickly and as accurately as possible to indicate whether the left, or right test probe corresponded to the average size of the preceding set of circles. Observers pressed "1" on the keyboard number pad to indicate the left circle corresponded to the mean circle size, and "2" to indicate the right circle. The location of each test circle was chosen at random on a trial-to-trial basis. Immediately following this response, observers were prompted to indicate the orientation of either the large (distributed attention condition) or small (focused attention condition) rectangle. Observers pressed "1" to indicate the rectangle was vertically oriented and "2" to indicate that it was horizontally oriented.

Observers completed 2 blocks of 300 trials. Each block corresponded to an attention condition (distributed or focused) and the order of blocks was randomized. Prior to each block, observers completed 20 practice trials to familiarize themselves with the task. Practice trials were not analyzed. An example of a trial sequence for Experiment 1 is displayed in Figure 3.1.

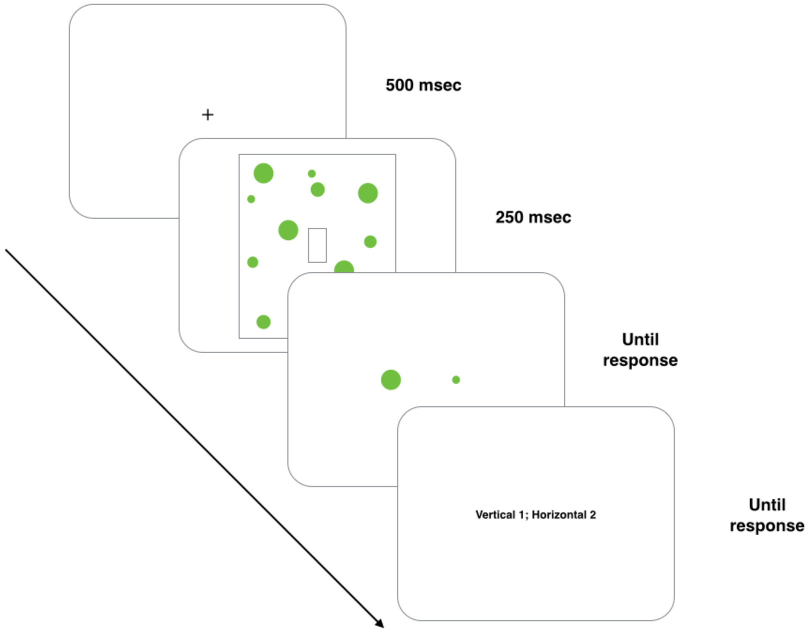


Figure 3.1. An example of a trial sequence in Experiment 1.

3.3.2 Results

3.3.2.1 Attention task accuracy. Mean accuracy for the attention task is displayed in Figure 3.2. Overall, accuracy was high (~ 87%), replicating the approximately 85% accuracy rate reported by Chong and Treisman (2005). There was no significant difference between distributed (M = .88; SD = 0.11) and focused (M = .85; SD = 0.09) attention conditions, suggesting that both tasks were equally difficult, $t(13) = 1.77$, $p >$ .101, Cohen's d = .48, a mean difference of .03 (SD = .07; 95% CI [-.01, .07]).

3.3.2.2 Mean discrimination task. Mean accuracy for the mean discrimination task is displayed in Figure 3.3a. Similar to the attention task, there was no statistically significant difference between groups, suggesting that the computation of mean size was equally difficult between distributed (M = .69; SD = .09) and focused attention conditions (M = .68; SD = 0.09), $t(13) = .281$, $p > .783$, Cohen's d = .07. This difference was less than .01 (SD = .08; 95% CI [-.04, .05]). Critically, performance in this task averaged 68% +- 7%. Given that we purposely manipulated the size of the test circles to replicate the 75% accuracy rate achieved by Chong and Treisman (2005), we were relatively close in obtaining our desired effect.

Mean RTs for each trial type are displayed in Figure 3.3b. Mean discrimination RTs were statistically significantly faster when completed concurrently with a distributed attention task (M = 1986.30; SD = 85.26) compared to a focused attention task (M = 2093.63; SD = 85.06), $t(13) = 2.27$, $p < .041$, Cohen's d = .61. The difference between these groups was approximately 107.33 msec (SD = 170.53; 95% CI [9.57, 205.09]).

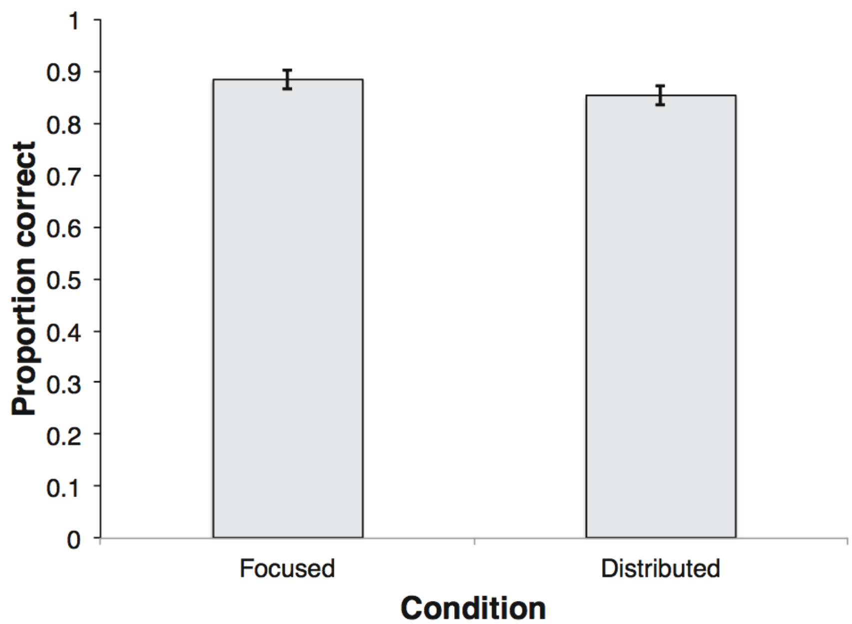


Figure 3.2. Mean accuracy for distributed and focused attention tasks in Experiment 1. The error bars represented here are the 95% within subject confidence intervals described by Loftus and Masson (1994).

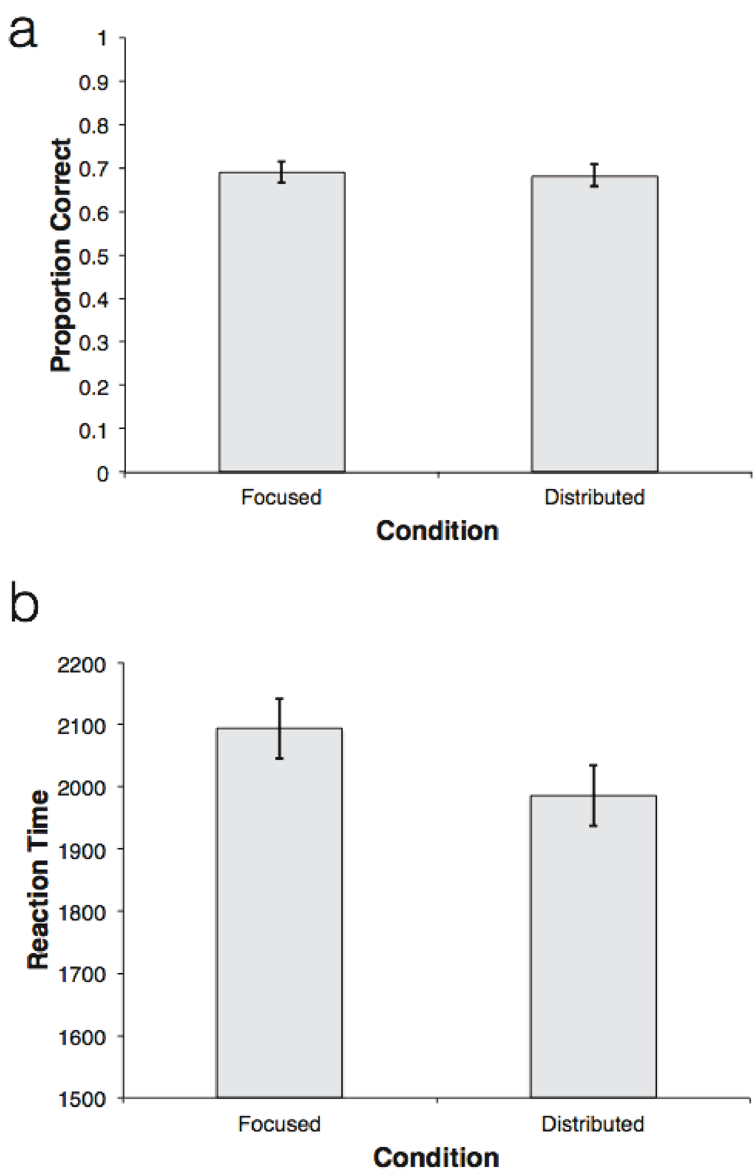


Figure 3.3. Mean discrimination results in Experiment 1. a) Mean discrimination task accuracy in distributed and focused attention conditions; b) Mean discrimination reaction times in distributed and focused attention conditions. The error bars represented here are the 95% within subject confidence intervals described by Loftus and Masson (1994);.

3.3.3 Discussion

The purpose of Experiment 1 was to replicate the previous finding that judgments about mean circle size are more easily combined with a distributed attention task compared to a focused attention task. Whereas Chong and Treisman (2005) indexed computation of mean circle size using threshold differences, we concentrated our analysis on RTs in order to obtain a baseline pattern of results with which the results of scene categorization performance in Experiments 2 - 4 could be compared. Critically, there was no statistically significant difference in accuracy between distributed and focused attention tasks, replicating Chong and Treisman's result that both tasks are equally difficult. Similarly, there was no difference in mean discrimination accuracy between attention conditions; an expected result given that we purposely made the diameter difference between the two test probes large enough to achieve 75% performance. The critical finding from Experiment 1 is that our observers were faster at judging mean circle size when they judged the orientation of a large bordering rectangle, corroborating Chong and Treisman's finding that the extraction of mean circle size is more easily combined with distributed rather than focused attention.

3.4 Experiment 2

Experiment 2 was designed to address which distribution of attention was more easily combined with the ability to detect the presence of an animal. The decision to index scene categorization using an animal detection task was based on previous studies that have used similar stimuli (Li et al., 2002; Rousselet et al., 2002; Thorpe et al., 1996; Walker et al., 2008). Experiment 2 was thus a replication of Experiment 1, with the exception that we replaced the mean discrimination task with an animal detection task. Furthermore, because the addition of secondary attention tasks affects animal detection sensitivity (Cohen et al., 2011), we wanted to obtain a baseline measure of performance; therefore, a control condition was included in which observers completed only the animal detection task.

3.4.1 Method

3.4.1.1 Observers. Ten naïve Concordia undergraduate students received partial course credit for their participation, or were paid $10 in monetary compensation. None of the observers participated in any of the other studies reported in this article, and all self-reported normal or corrected-to-normal vision.

3.4.1.2 Stimuli and apparatus. Stimuli and apparatus were the same as in Experiment 1 with the following exceptions. Stimuli were 1000 pictures downloaded from the Corel image database (Corel, 1996) organized into a target present and target absent categories[1]. Images were gray scaled and were presented on a gray background (RBG values [128 128 128]; luminance of 52 cd/m2). The target category contained

[1] For image examples please see the Corel stock photo library (copyright 1996 by Corel). Permission for reprints is granted for published articles only.

images of animals that included birds, insects, reptiles, mammals, and fish. The distractor category contained images of foods, fruits, plants and vehicles. The stimuli were 256 X 256 pixels. When viewed at a distance of 54 cm the images subtended a visual angle of 10.5° X 10.5°.

3.4.1.3 Procedure. Each trial began with a fixation cross located in the middle of the screen presented for 500 msec. The primary purpose of the present study was to examine how different attentional distributions affected concurrent rapid scene categorization behaviour. Thus, we were particularly concerned with ensuring that our observers were able to perceive our scene stimuli, while also limiting processing time. Immediately following presentation of the images, we therefore presented a white noise mask (amplitude spectrum = 0) for a period of 64 msec. Hansen and Loschky (2013) showed that masks constructed in this way result in approximately 80% scene categorization performance. As in Experiment 1, a large rectangle bordered the image and a small rectangle was located at the centre of the image. On 50% of trials, the image contained an animal; the image was a distractor on the other 50% of trials. We instructed observers to as quickly and as accurately as possible to indicate whether the image presented contained an animal (or not). Observers pressed the “1” key on the keyboard number pad to indicate the presence of an animal, and the “2” key to indicate the absence of an animal. Immediately following this response, observers were prompted to indicate the orientation of either the large (distributed attention condition), or small (focused attention condition) rectangle. Observers pressed the “1” key to indicate that the rectangle was vertically oriented and the “2” key to indicate that it was horizontally oriented. In a control condition, observers completed only the animal detection task. If observers did not respond to the animal detection task within 4000 msec, then the trial was discarded. This removed less than 1% of total trials.

Observers completed 3 blocks of 300 trials. Each block corresponded to a different experimental condition (distributed, focused, or control) and the order of blocks was chosen at random. Prior to the start of each block, observers completed 20 practice trials in order to familiarize themselves with the task. Practice trials were not analyzed. An example of a trial sequence for Experiment 2 is displayed in Figure 3.4.

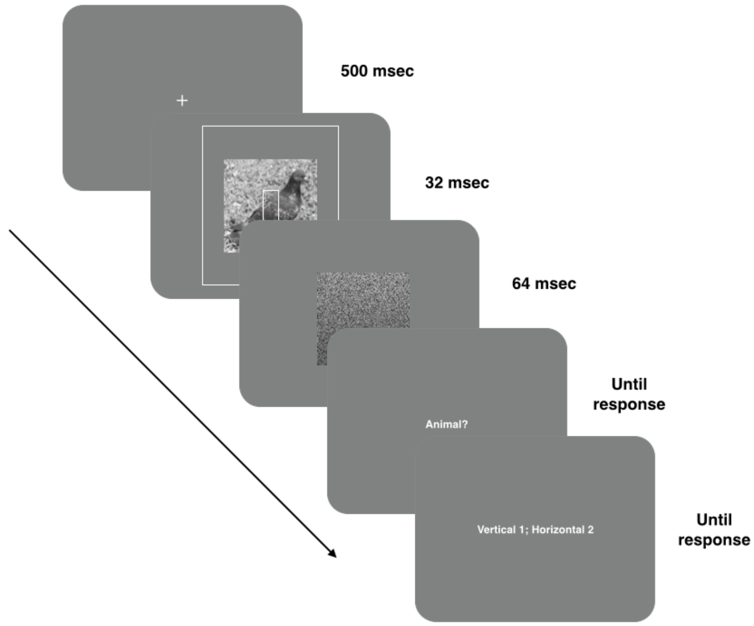


Figure 3.4 An example of a trial sequence in Experiment 2.

3.4.2 Results

3.4.2.1 Attention task accuracy. Mean accuracy for the attention task is displayed in Figure 3.5. As in Experiment 1, accuracy was overall high (~ 90%) and there was no significant difference between distributed (M = .93; SD = 0.04) and focused (M = .89; SD = 0.03) attention tasks, $t(9) = 1.93$, $p > .09$, Cohen's $d = .06$. This difference was approximately .04 (SD = .05; 95% CI [-.01, .09]). It is interesting to note that both the mean difference and effect size were similar to those reported in Experiment 1, suggesting that attention task difficulty is similar between experiments.

3.4.2.2 Animal detection sensitivity. We compared animal detection performance using the sensitivity measure, d prime (d'; Figure 3.6a). Overall, d' values were high (d' = 3.32; hit rate = .95%; false alarm rate = .09%). We entered these means into a one-way repeated measures ANOVA. There was no statistically significant main effect, suggesting that the addition of both distributed and focused attention tasks did not affect sensitivity to the presence of an animal, $F(2, 18) = 2.22$, $p > .138$, $\eta^2 = .31$. d' statistics are displayed in Table 3.1.

3.4.2.3 Animal detection task RTs*.* As in Experiment 1, we computed group mean RTs for each trial type. Trial type varied according to condition and whether the target was present, or absent. These group means are displayed in Figure 3.6b. We entered these means into a 2 (target present or absent) X 3 (experiment condition: distributed, focused, or control) repeated measures ANOVA. There were significant main effects of target, $F(1,9) = 9.14$, $p < .013$, $\eta^2 = .51$, and condition, $F(2, 18) = 51.84$, $p < .001$, $\eta 2 = .001$, η^2 = .86. The target X condition interaction was also significant, $F(2,18) = 3.72$, $p < .044$, $\eta 2 = .29$. Looking at the figure, it appears that the main effects were driven by the fact that RTs were fastest on single task trials and on target present displays. We confirmed this interpretation by computing 1) the contrast comparing RTs between single task trials (M = 736.19; SD = 105.75) and dual task trials (M = 1200.14; SD = 55.46) (Ψ_1); and 2) the contrast comparing RTs between target present trials (M = 1023.40; SD = 20.65) and target absent trials (M = 1067.72; SD = 21.67) (Ψ_2). Both these contrasts were statistically significant.

Because Experiment 2 was designed to assess which attentional distribution was more easily combined with the detection of an animal, we were primarily interested in the planned contrasts comparing performance between distributed and focused attention conditions for both target present (Ψ_3) and target absent (Ψ_4) trials. Consistent with the significant target X condition interaction, animal detection was faster when combined with a distributed attention task (M = 1104.71; SD = 71.26) compared to a focused attention task (M = 1218.47; SD = 109.20) on target present trials. There was no significant difference between distributed (M = 1202.71; SD = 68.8) and focused attention tasks (M = 1275.34; SD = 104.49) on target absent trials. Corroborating this result, the effect size for target present (η^2 = .43) was larger than the effect size for target absent displays (η^2 = .23), although this difference is relatively small. Reaction time statistics and contrast analyses are displayed in Table 3.2.

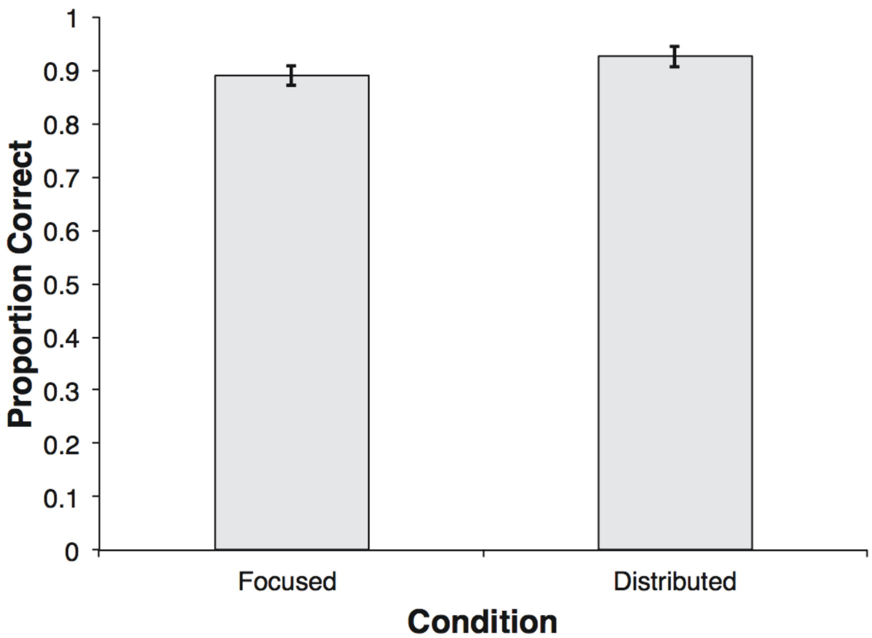


Figure 3.5. Mean accuracy for distributed and focused attention tasks in Experiment 2. The error bars represented here are the 95% within subject confidence intervals described by Loftus and Masson (1994).

Table 3.1. *d prime* statistics for each condition in Experiment 2.

	Condition								
	Control			Focused			Distributed		
Statistic	***M***	***SD***	**95% CI**	***M***	***SD***	**95% CI**	***M***	***SD***	**95% CI**
***d* prime**	3.56	0.55	[3.17, 3.96]	3.35	0.82	[2.76, 3.94]	3.04	.68	[2.55, 3.53]
Hit rate	0.97	0.03	[.94, .99]	0.96	0.03	[.93, .98]	0.94	.04	[.92, .97]
False alarm rate	0.07	0.06	[.03, .12]	0.11	0.11	[.03, .19]	0.09	.05	[.07, .13]

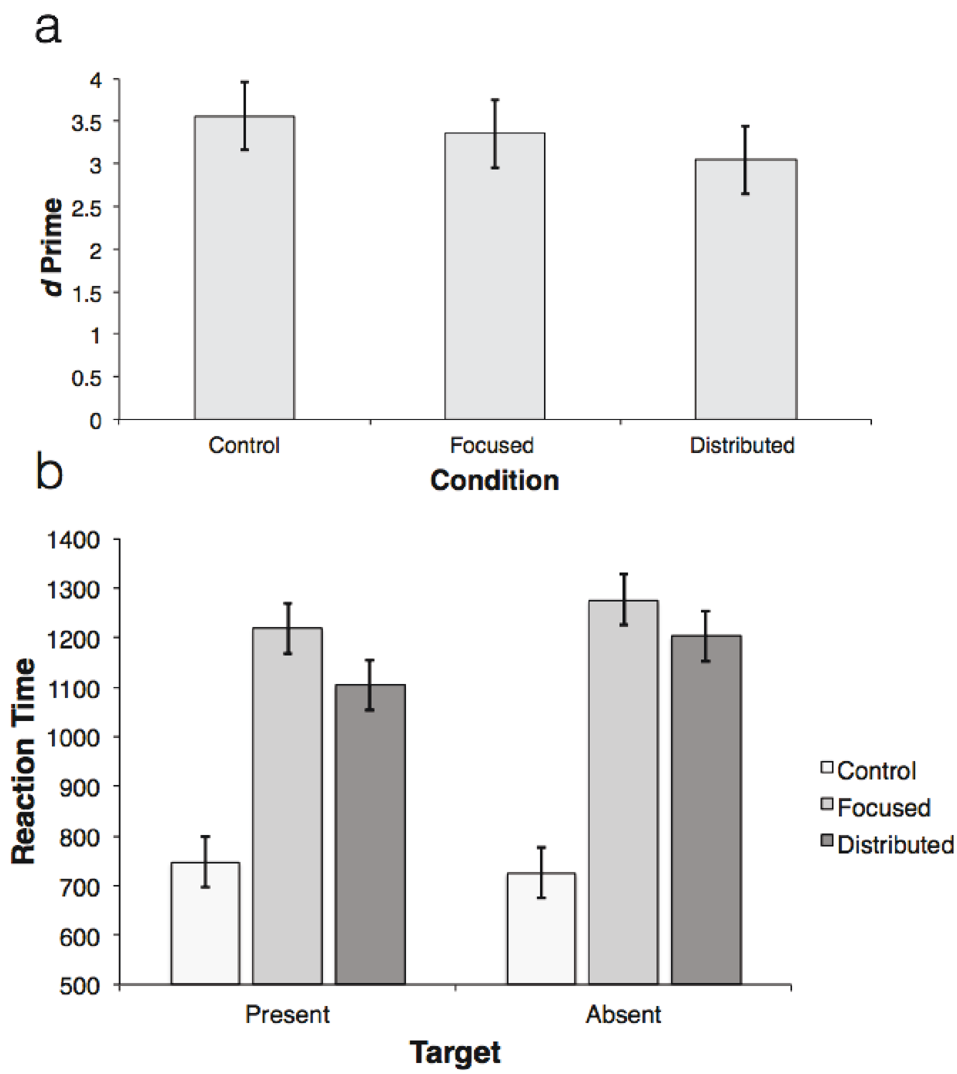


Figure 3.6. Animal detection results in Experiment 2. a) Mean *d prime* values in each condition; b) Animal detection reaction times in each condition on target present and target absent displays. The error bars represented here are the 95% within subject confidence intervals described by Loftus and Masson (1994).

Table 3.2. Reaction time (msec) statistics for each condition in Experiment 2. Reaction time mean difference contrasts in Experiment 2

	Target Present			Target Absent		
Condition	***M***	***SD***	**95% CI**	***M***	***SD***	**95% CI**
Control	747.28	117.28	[663.12, 831.45]	725.10	125.10	[635.29, 814.93]
Focused	1218.47	109.20	[1140.35, 1296.59]	1275.34	104.49	[1200.64, 1350.04]
Distributed	1104.44	71.26	[1050.97, 1157.90]	1202.72	68.80	[1151.09, 1254.33]

Contrasts

Contrast	***df***	***F***	***p***	M_D	SE (M_D)	95% CI (M_D)	$\eta 2$
Ψ_1	(1, 9)	70.01	< .001	464.05	52.61	[345.03, 583.06]	.88
Ψ_2	(1, 9)	9.41	< .013	44.31	13.57	[13.32, 75.31]	.51
Ψ_3	(1, 9)	6.83	< .028	114.03	41.40	[20.38, 207.68]	.43
Ψ_4	(1, 9)	2.28	> .143	72.62	42.92	[-24.46, 169.71]	.23

Ψ1: RT comparison between single and dual-task trials.
Ψ2: RT comparison between target present and target absent trials.
Ψ3: RT comparison between distributed and focused attention conditions on target present displays.
Ψ4: RT comparison between distributed and focused attention conditions on target absent displays.

3.4.3 Discussion

The aim of Experiment 2 was to examine whether distributed or focused attention was more compatible with the rapid detection of an animal. Overall, attention task performance replicated the results of Experiment 1, suggesting that the difficulty between distributed and focused attention tasks was approximately equal. Furthermore, animal detection sensitivity was high, suggesting a weak effect of masking. Nevertheless, this result is not completely unexpected given that Hansen and Loschky (2013) reported an 80% scene categorization detection rate using the same mask. Furthermore, this high animal detection performance is consistent with previous studies that reported 90% animal detection rates, although with unmasked presentations (see, e.g., Walker et al., 2008).

The addition of a simultaneous attention task slowed animal detection responses, replicating the finding that the addition of a secondary task slows responses to a primary scene categorization task (Walther & Fei Fei, 2007). However, this slowing did not have an effect on animal detection sensitivity. Furthermore, although RTs were slower on dual task trials, observers were faster to respond to the presence of an animal when simultaneously completing a distributed attention task, compared to a focused attention task. The present results are thus consistent with both Experiment 1 and Evans and Treisman's (2005) suggestion that distributed attention facilitates the rapid extraction of a scene's summary statistics that can be used to categorize a scene. However, because Experiment 2 used an animal detection task, the present results are limited to comparisons with previous studies that have used similar paradigms (Li et al., 2002; 2005; Rousselet et al., 2002; Walker et al., 2010; Thorpe et al., 1996). Thus, in

Experiment 3, we examined whether the results of Experiments 1 and 2 extend to the categorization of more complex scenes.

3.5 Experiment 3

Experiment 3 was designed to investigate whether the results of Experiment 2 extend to more complex natural scene categorization tasks. Whereas scene categorization performance was measured using an animal detection task in Experiment 2, observers classified natural scenes according to their basic level in Experiment 3.

3.5.1 Method

3.5.1.1 Observers. Twenty-two Concordia university undergraduate students participated in this experiment in return for partial course credit. None of the observers participated in the other studies reported in this article, and self-reported normal or corrected-to-normal vision. Data from two participants was discarded because they failed to achieve chance performance in any of the conditions.

3.5.1.2 Stimuli, apparatus, and procedure. Stimuli, apparatus, and procedure were the same as in Experiment 2 with the following exceptions. Scenes were over 500 images of beaches, rivers, mountains, forests, and deserts taken from the Corel image database (Corel, 1996) and the Sun image database (Xiao, Hayes, Ehinger, Oliva, & Torralba, 2010). All images were gray scaled. All scenes were the same size as Experiment 2 and presented on the same gray background. Mathwork's Matlab (ver. 2011b) with the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997; Kleiner et al, 2007) controlled all timing and data recording operations.

Each trial began with a fixation cross presented in the middle of the screen presented for 500 msec, immediately followed by the presentation of a scene for 32 msec. As in Experiment 1, a white noise masked presented for 64 msec was used to mask the scene. As in Experiments 1 and 2, a large rectangle bordered the image and a small rectangle was located centrally. Immediately following offset of the mask, observers were presented with a cue word in the middle of the screen presented until response. On 50% of trials, the cue word matched the target category; on the other 50% of trials, the cue word matched a distractor category chosen at random. We instructed observers to indicate whether the cue word matched the previously presented image as quickly and as accurately as possible. Observers pressed the "1" key on the keyboard number pad to indicate that the cue matched the image and the "2" key to indicate that it did not. Immediately following this response, we then prompted observers to indicate the orientation of either the large (distributed attention condition), or small (focused attention condition) rectangle. Similar to Experiments 1 and 2, observers pressed the "1" key to indicate that the rectangle was vertically oriented and the "2" key to indicate that it was horizontally oriented. In a control condition, observers completed only the scene categorization task. If observers did not respond to the cue word within 4000 msec, then the trial was discarded. This eliminated less than 1% of total trials.

Similar to Experiment 2, observers completed 3 blocks (distributed, focused, and control) of 300 trials and the order of the blocks was chosen at random. The target category and the probe word varied from trial-to-trial. Prior to start of each block, observers complete 20 practice trials. Practice trials were not analyzed. An example of a trial sequence in Experiment 3 is displayed in Figure 3.7.

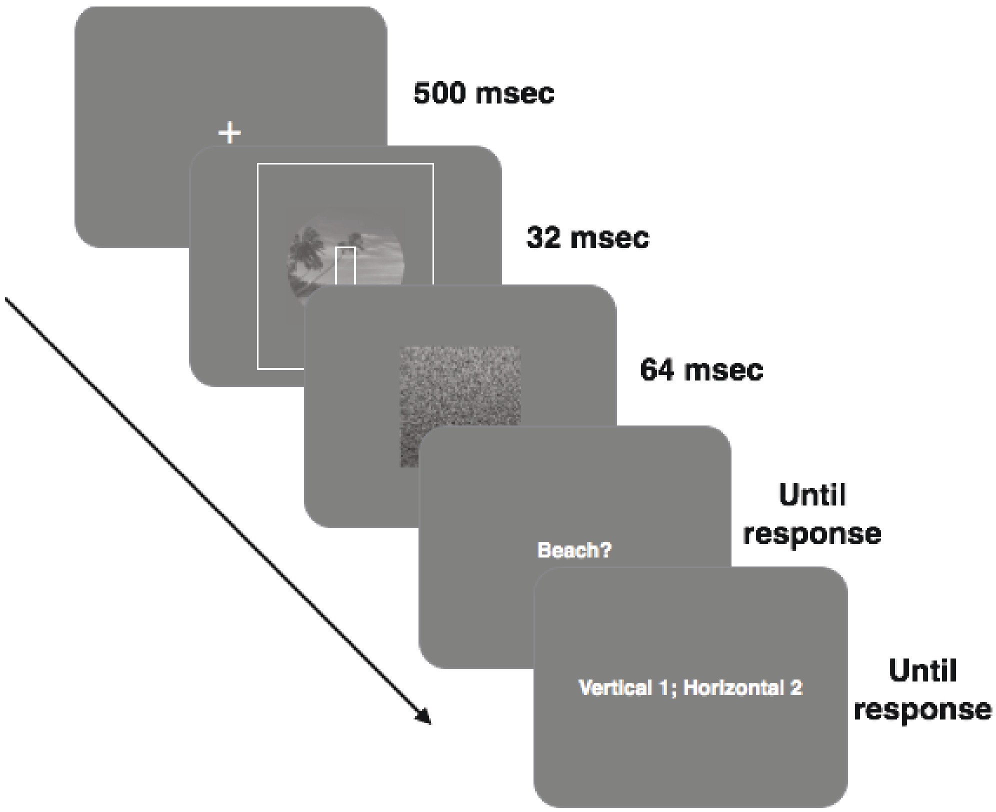


Figure 3.7. An example of a trial sequence in Experiment 3.

3.5.2 Results

3.5.2.1 Attention task accuracy. Mean accuracy is displayed in Figure 3.8. Overall, accuracy was high (82%), replicating the high performance in the previous experiments. Similar to Experiments 1 and 2, there was no statistically significant difference in accuracy between distributed (M = .82; SD = 0.11) and focused attention tasks (M = .83; SD = 0.16), $t(19) = .727$, $p > .354$, Cohen's d = .08. This difference was approximately .01 (SD = .13; 95% CI [-.05, .07]). Both the mean difference and effect size were similar to Experiments 1 and 2, corroborating our previous suggestion that attention task difficulty is approximately equal between experiments.

3.5.2.2 Scene categorization sensitivity. As in Experiment 2, scene categorization accuracy was measured using the sensitivity measure, d' (Figure 3.9a) and analyzed using a one-way repeated measures ANOVA. Overall, sensitivity was high (d' = 1.88; hit rate = .86%; false alarm rate = .22%). Consistent with Experiments 1 and 2, there was no statistically significant main effect, suggesting that the addition of simultaneous distributed and focused attention tasks did not affect scene categorization sensitivity, $F(2,38) = 2.19$, $p > .126$, $\eta^2 = .26$. Furthermore, the effect size was similar to the $\eta^2 = .31$ value reported in Experiment 2, suggesting that the effects of simultaneous attention tasks on scene sensitivity were the same between animal detection and basic level categorization tasks. d' statistics are displayed in Table 3.3.

3.5.2.2 Scene categorization RTs. Scene categorization results are displayed in Figure 3.9b. Similar to the previous experiments, we entered mean RTs into a 2 (target: present or absent) X 3 (condition: distributed, focused, or control) repeated measures ANOVA. There were significant main effects of target, $F(1,19) = 39.41$, $p < .001$, η^2 =

.68, and condition, $F(2,38) = 47.46$, $p < .001$, $\eta^2 = .71$. The target X condition interaction was not significant, $F(2, 38) = 2.28$, $p > .118$, $\eta^2 = .11$. Similar to Experiment 2, the main effects were driven by the fact that target absent trials ($M = 1178.64$; $SD = 72.01$) were slower than target present trials ($M = 970.87$; $SD = 72.13$) (Ψ_1), and by the fact that single task trials ($M = 825.96$; $SD = 142.73$) were faster than dual task trials ($M = 1199.14$; $SD = 71.36$) (Ψ_2). Furthermore, the effect sizes associated with the compassions between target present and target absent trials ($\eta^2 = .68$) and between single and dual task trials ($\eta^2 = .76$) were consistent with the effect sizes reported in Experiment 2 ($\eta^2 = .51$ and $\eta^2 = .88$), suggesting that the size of the effects are approximately equal between experiments.

As in Experiment 2, we were particularly interested in the planned contrast comparing RTs between focused and distributed attention conditions for both target present and target absent displays (Ψ_3). However, because there was no significant interaction between condition and target presence, we collapsed across target present and target absent trials. There was no statistically significant difference between focused ($M = 1172.17$; $SD = 90.31$) and distributed attention conditions ($M = 1226.12$; $SD = 96.06$) (Ψ_3). What's more, the associated effect size ($\eta^2 = .17$) was relatively smaller than the reported effect sizes for both target present ($\eta^2 = .51$) and target absent trials ($\eta^2 = .43$) in Experiment 2. This suggests that the effects of distributed attention were greater on the animal detection task in Experiment 2 compared to the basic level scene categorization task in Experiment 3**.** Reaction time statistics and contrast analyses are displayed in Table 3.4.

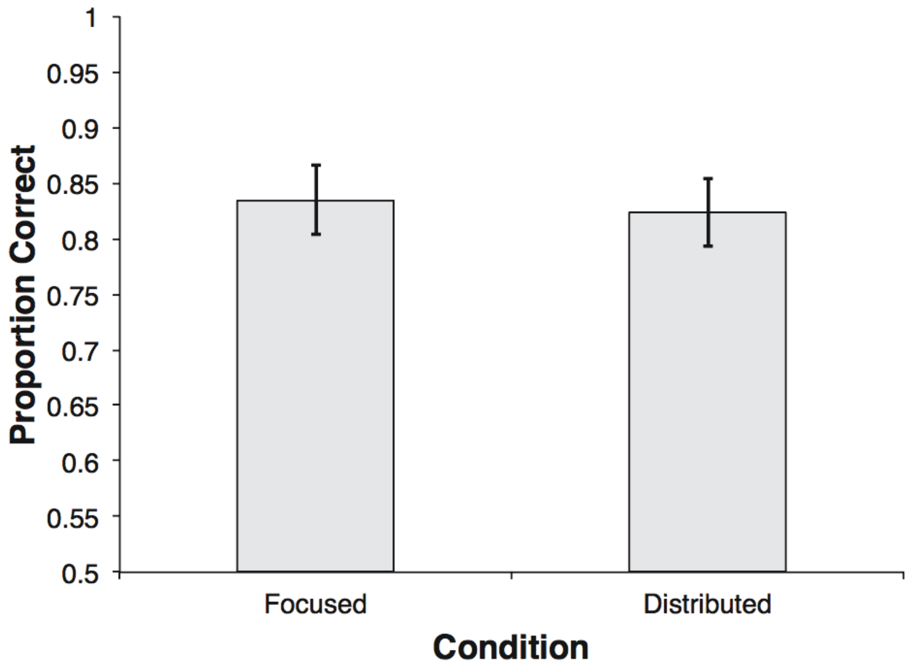


Figure 3.8. Mean accuracy for distributed and focused attention tasks in Experiment 3. The error bars represented here are the 95% within subject confidence intervals described by Loftus and Masson (1994).

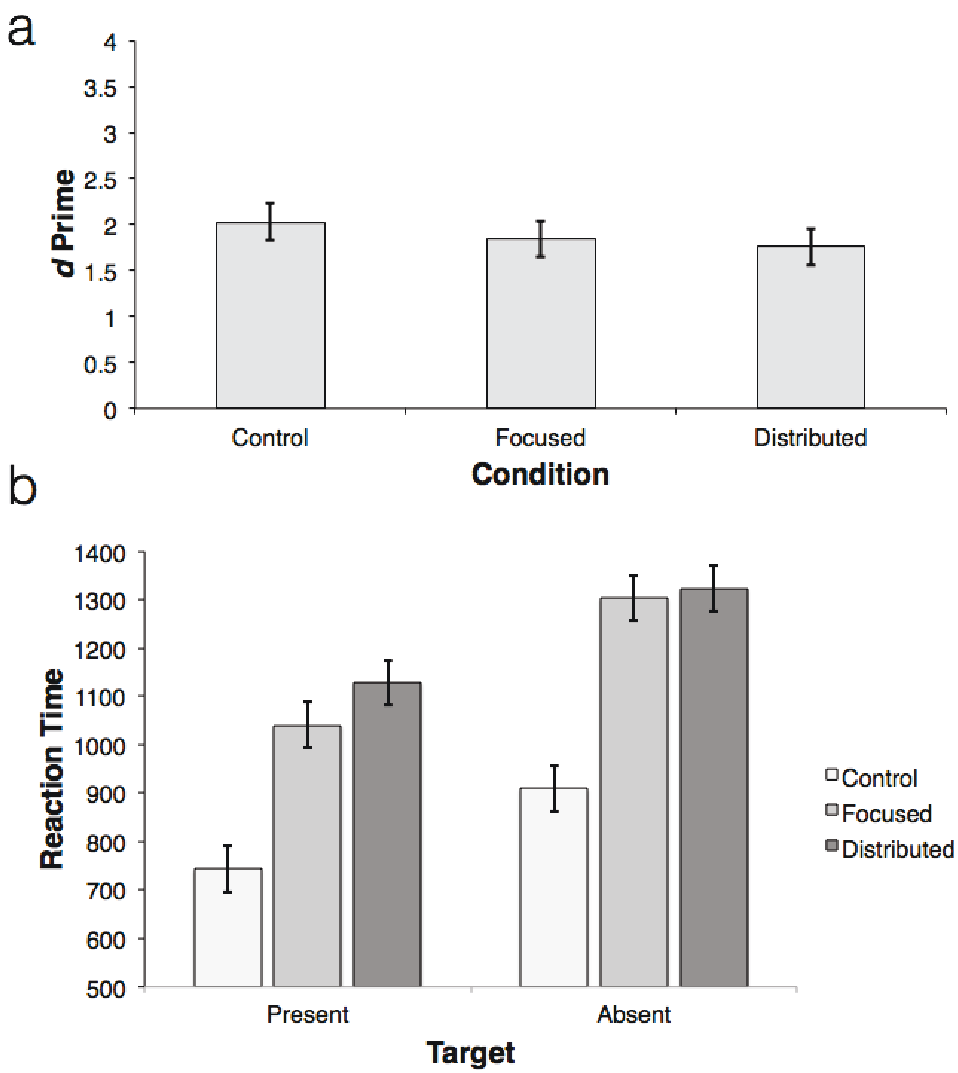


Figure 3.9. Basic level scene categorization results in Experiment 3. a) Mean *d prime* values for each condition; b) Scene categorization reaction times for each condition on target present and target absent trials. The error bars represented here are the 95% within subject confidence intervals described by Loftus and Masson (1994).

Table 3.3. *d prime statistics for each condition in Experiment 3.*

	Condition								
	Control			Focused			Distributed		
Statistic	***M***	***SD***	**95% CI**	***M***	***SD***	**95% CI**	***M***	***SD***	**95% CI**
***d* prime**	2.03	0.26	[1.20, 2.86]	1.84	0.03	[0.95, 2.73]	1.76	0.41	[0.65, 2.86]
Hit rate	.83	0.12	[.72, .95]	.87	0.08	[.81, .93]	.86	0.09	[.82, .90]
False alarm rate	.21	0.12	[-.03, .46]	.21	0.08	[.06, .36]	.23	0.06	[.13, .34]

Table 3.4. Reaction time (msec) statistics in Experiment 3. Reaction time mean difference contrasts in Experiment 3

	Target Present			Target Absent		
Condition	***M***	***SD***	**95% CI**	***M***	***SD***	**95% CI**
Control	743.02	175.06	[661.08, 824.95]	908.89	145.09	[840.99, 976.81]
Focused	1040.77	123.11	[983.16, 1098.39]	1303.57	168.69	[1224.61, 1382.52]
Distributed	1128.81	117.68	[1073.74, 1183.89]	1323.43	135.18	[1260.18, 1386.71]

Contrasts

Contrast	***df***	***F***	***p***	M_D	$SE(M_D)$	**95% CI**(M_D)	$\eta 2$
Ψ_1	(1, 19)	39.41	< .001	207.77	32.26	[140.25, 275.29]	.68
Ψ_2	(1, 19)	57.73	< .001	373.18	47.87	[272.99, 473.39]	.76
Ψ_3	(1, 19)	3.84	> .072	53.95	26.82	[-2.19, 110.10]	.17

Ψ1: RT comparison between target present and target absent trials.
Ψ2: RT comparison between single and dual-task trials.
Ψ3: RT comparison between distributed and focused attention conditions.

3.5.3 Discussion

The purpose of Experiment 3 was to examine the effects of focused and distributed attention on basic level scene categorization. Similar to Experiments 1 and 2, attention task performance did not differ statistically between distributed and focused attention tasks, suggesting that they were equally difficult. Consistent with Experiment 2, the presence of simultaneous attention tasks did not affect sensitivity to scene categorization, but did slow responses. In contrast, however, whereas responses to the presence of an animal were faster when combined with a distributed attention task in Experiment 2, there was no difference in RTs between distributed and focused attention conditions in response to the basic level of a scene in the present experiment.

One explanation for this discrepancy is that we failed to replicate Chong and Treisman's result in Experiment 3; however, this is unlikely as we successfully replicated their study in Experiments 1 and 2. An alternative explanation is that our animal detection task demands differed from our scene categorization task demands. Our animal detection task required observers to report the presence of a target that remained constant throughout the entire experiment. In contrast, the target scenes in our basic level scene categorization task varied from trial-to-trial, and could have been any one of a possible five scenes (e.g., beaches, rivers, mountains, forests, or deserts). Evans, Horowitz, and Wolfe (2011) provided evidence to suggest that when the target category is unknown, interference effects are more likely to emerge than when it is known. Evans and colleagues (Experiment 1) asked participants to indicate whether a target image was present in RSVP stream. The target category varied from trial-to-trial and could have been any one of nine different categories. In one block of trials, the target category was pre-cued. In another block of trials, the target category was post-cued. Although *d'* values were above 1.5 in both conditions, they were higher in the pre-cued condition than in the post-cued condition, suggesting that there was a cost associated with holding more than one target category in memory.

In Experiment 4, we directly tested this hypothesis by modifying the scene categorization task in Experiment 3, such that the task demands were similar to the task demands associated with the animal detection task in Experiment 2. Specifically, observers were asked to complete a scene categorization task based on the superordinate natural/manmade distinction. Thus, similar to the animal detection task, observers were asked to base classification on the superordinate level, and were required to hold only a single target category in memory.

3.6 Experiment 4

3.6.1 Method

3.6.1.1 Observers. Ten Concordia University undergraduate students participated in this experiment in return for partial course credit, or were paid $10 in monetary compensation. None of the observers participated in the other studies reported in this article, and self-reported normal or corrected-to-normal vision.

3.6.1.2 Stimulus, apparatus, and procedure. Stimuli, apparatus, and procedure were the same as in Experiments 2 and 3 with the following exceptions. Natural scene stimuli consisted of over 500 each of beach, desert, forest, and mountain scenes. Manmade stimuli consisted of over 500 each of city, highway, and living room scenes. Both manmade and natural scenes were taken from the Corel (Corel, 1996) image database. The attention task was completed as described in the previous experiments. The scene categorization task was completed as follows. We instructed observers to indicate whether the image presented corresponded to a natural image. Observers pressed the "1" key on the keyboard number pad to indicate that the image corresponded to a natural image, and the "2" key to indicate that it did not. Similar to the previous experiments, observers completed 3 blocks of 300 trials, each block corresponding to a different experimental condition. Less than 1% of trials were removed because of a failure to respond within 4000 msec. An example of a trial sequence in Experiment 4 is displayed in Figure 3.10.

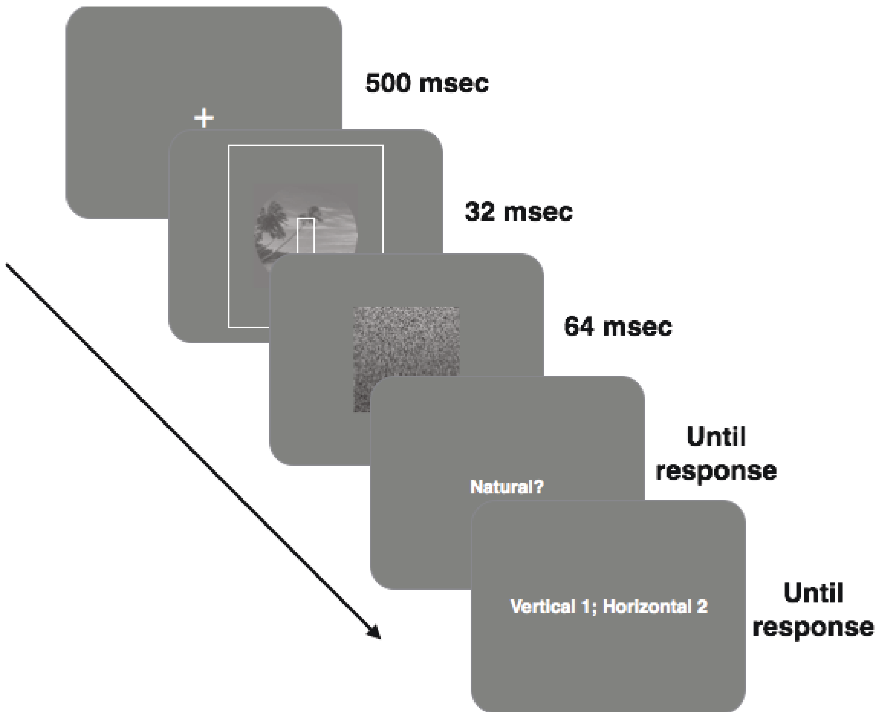


Figure 3.10. An example of a trial sequence in Experiment 4.

3.6.2 Results

3.6.2.1 Attention task accuracy. Attention task results are displayed in Figure 3.11. As in the previous experiments, accuracy was overall high (~ 92%). However, and in contrast to the previous experiments, performance on the distributed attention task (M = .94; SD = 0.03) was statistically significantly higher than the focused attention task (M = .90; SD = 0.04), $t(9) = 4.25$, $p < .002$, Cohen's $d = 1.3$. This was a difference of approximately .04 (SD = .03; 95% CI [.03, .05]). It is interesting to note that although the absolute difference between distributed and focused attention tasks was similar to the previous experiments, the effect size was relatively higher than Experiments 1 - 3, suggesting that difference in accuracy between distributed and focused attention tasks is greatest in Experiment 4.

3.6.2.2 Scene categorization sensitivity. The d' values for the scene categorization task are displayed in Figure 3.12a**.** As in previous experiments, sensitivity was overall high (d' = 2.93; hit rate = .90%; false alarm rate = .09%). We entered these d prime values into a one-way repeated measures ANOVA. Consistent with the previous experiments, this analysis was not statistically significant, $F(2,18) = 2{,}54$, $p > .107$, η^2 = .22. Furthermore, the effect size in Experiment 4 was relatively small and similar to the values reported in previous experiments, corroborating the finding that the addition of simultaneous attention tasks does not affect sensitivity to scene categorization tasks. d' statistics are displayed in Table 3.5.

3.6.2.3 Scene categorization RTs. Mean scene categorization RTs are displayed in Figure 3.12b. As with previous experiments we analyzed these means using a 2 (target: present or absent) X 3 (condition: distributed, focused, or control) repeated measures ANOVA. There was a significant main effect of condition, $F(2,18) = 36.55$, $p < .001$, $\eta^2 = .81$. Neither the main effect of target nor the condition X target interaction was significant, $F(1,9) = .06$, $p > .814$, $\eta^2 < 01$, and $F(2, 38) = 1.58$, $p > .201$, $\eta^2 = .16$. In contrast to previous experiments, there was no statistically significant difference between target present ($M = 939.51$; $SD = 59.88$) and target absent trials ($M = 949.20$; $SD = 59.01$) a difference of 9.69 msec ($SD = 119.76$; 95% CI [-77.64, 97.02]).

As in Experiments 2 and 3, we computed the planned contrast comparing RTs between 1) single and dual task trials (Ψ_1), and 2) between focused and distributed attention trials, collapsing across target presence (Ψ_2). Single task trials ($M = 635.71$; $SD = 121.23$) were statistically significantly faster than dual-task trials ($M = 1098.89$; $SD = 60.61$). In addition, the effect size ($\eta^2 = .87$) was similar to Experiments 2 than 3, corroborating our suggestion that the effect of adding a secondary task on scene categorization performance is equal between experiments. Furthermore, observers were faster to respond on distributed attention trials ($M = 973.04$; $SD = 110.15$) than focused attention trials ($M = 1224.33$; $SD = 127.02$) corroborating the result from Experiment 2. Consistent with this result, the reported effect size ($\eta^2 = .61$) was more consistent with the effect size in Experiment 2 ($\eta^2 = .43$; target absent $\eta^2 = .23$) than 3 ($\eta^2 = .17$). This suggests that the effects of distributed attention on superordinate level scene categorization are more similar with the effects on distributed attention on animal detection (on target present trials) than the effects of distributed attention on basic level categorization. Reaction time statistics and contrast analyses are displayed in Table 3.6.

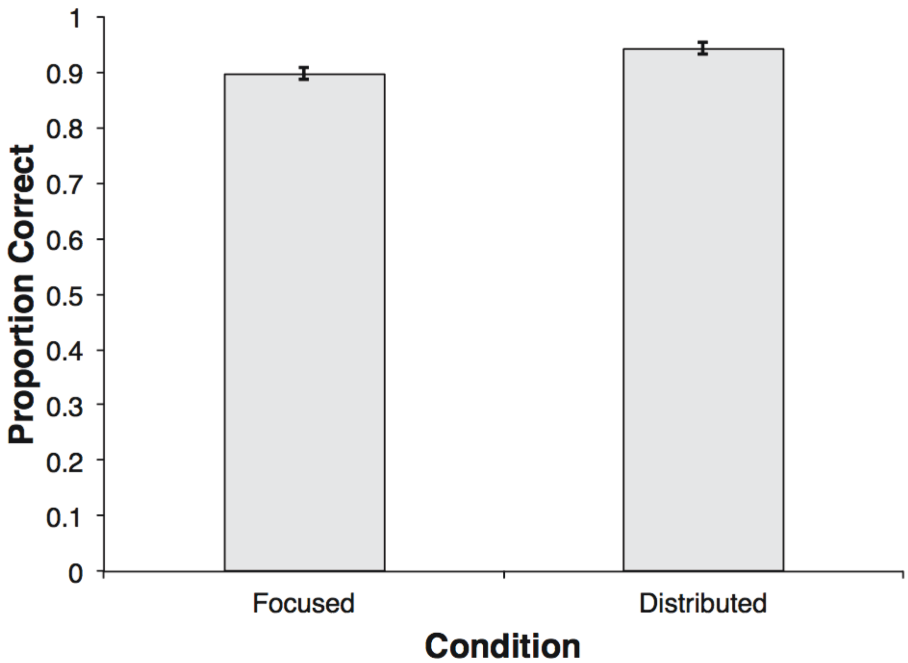


Figure 3.11. Mean accuracy for distributed and focused attention tasks in Experiment 4. The error bars represented here are the 95% within subject confidence intervals described by Loftus and Masson (1994).

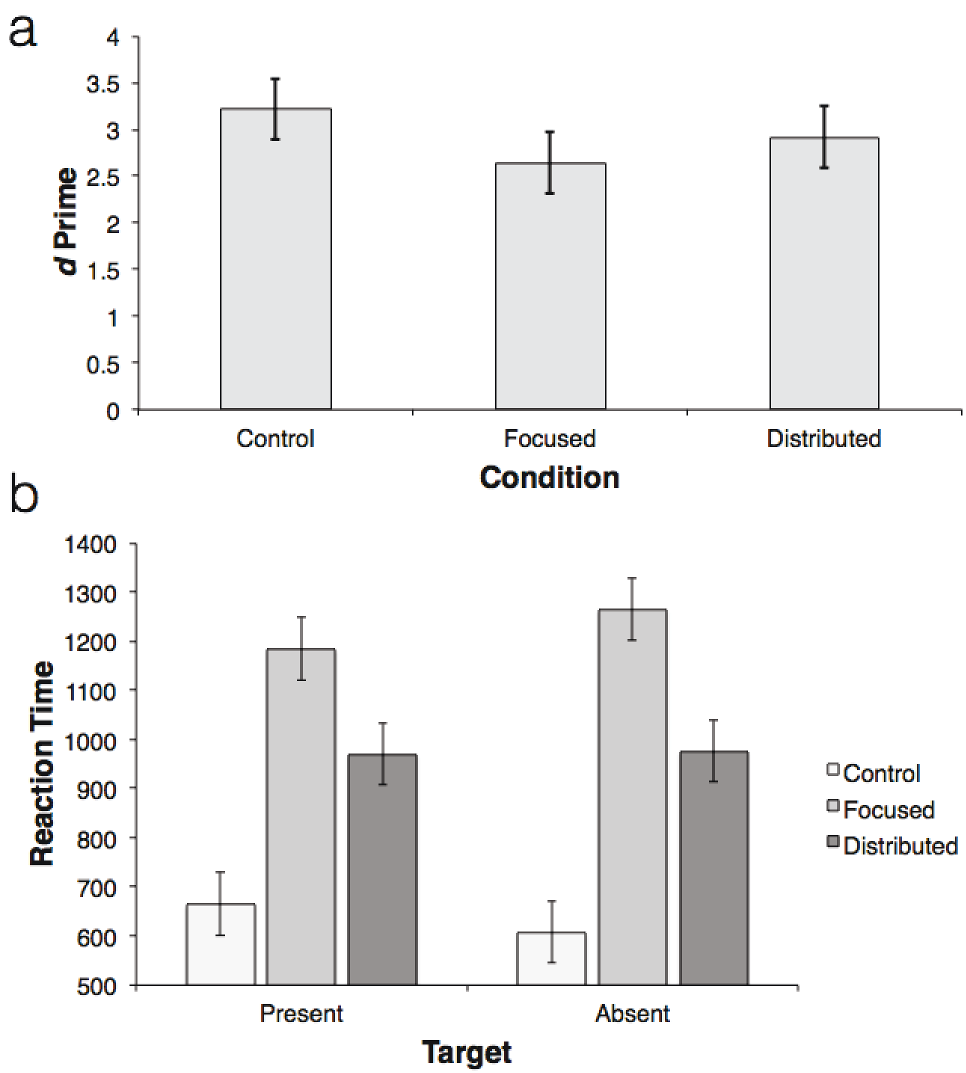


Figure 3.12. Superordinate level categorization results in Experiment 4. A) Mean *d prime* values in each condition; b) Scene categorization reaction times for each condition on target present and target absent displays. The error bars represented here are the 95% within subject confidence intervals described by Loftus and Masson (1994).

Table 3.5. *d prime* statistics for each condition in Experiment 4.

	Condition								
	Control			Focused			Distributed		
Statistic	***M***	***SD***	**95% CI**	***M***	***SD***	**95% CI**	***M***	***SD***	**95% CI**
***d* prime**	3.22	0.64	[2.81, 3.63]	2.64	0.82	[2.41, 2.86]	2.92	0.49	[2.56, 3.27]
Hit rate	.92	.06	[.87, .96]	.89	.02	[.88, .90]	.88	.06	[.84, .93]
False alarm rate	.05	.09	[-.02, .11]	.13	.06	[.09, .18]	.11	.06	[.06, .15]

Table 3.6. Reaction Time (msec) statistics in Experiment 4. Reaction time mean difference contrasts in Experiment 4

	Target Present			Target Absent		
Condition	***M***	***SD***	**95% CI**	***M***	***SD***	**95% CI**
Control	664.23	120.29	[576.51, 751.95]	607.19	132.41	[510.63, 703.75]
Focused	1184.13	166.25	[1062.89, 1305.36]	1264.53	206.99	[1113.58, 1415.47]
Distributed	970.19	129.28	[875.92, 1064.46]	975.89	111.78	[894.39, 1057.41]

Contrasts

Contrast	***df***	***F***	***p***	M_D	$SE(M_D)$	95% CI (M_D)	$\eta 2$
Ψ_1	(1, 9)	58.34	< .001	462.98	57.50	[330.37, 595.58]	.87
Ψ_2	(1, 9)	13.58	< .005	251.29	64.68	[102.13, 400.45]	.61

Ψ1: RT comparison between single and dual task trials.
Ψ2: RT comparison between distributed and focused attention conditions.

3.6.3 Discussion

Experiment 4 investigated whether distributed or focused attention was more compatible with superordinate level scene categorization. In contrast to previous experiments, accuracy on the distributed attention task was statistically significantly higher compared to the focused attention task. However, given that performance exceeded 90% in both conditions, it is unlikely that this difference affected scene categorization performance in a meaningful way. Consistent with the previous experiments, the addition of a simultaneous attention task did not affect scene categorization sensitivity, but did slow categorization RTs. In contrast to Experiment 3, and consistent with the results of Experiments 1 and 2, superordinate level categorization was faster when combined with a distributed attention task, compared to a focused attention task. However, whereas animal detection was faster when combined with a distributed attention task on target present trials (Experiment 2), superordinate level categorization was faster when combined with a distributed attention task on both target present and target absent trials.

3.7 General Discussion

Evans and Treisman (2005) suggested that distributed attention facilitates the extraction of summary statistics, which in turn, can be used to categorize a scene information. However, empirical investigations in support of this hypothesis have been limited to studies investigating the effects of attention on perceptual averaging (e.g., mean size of a set of circles; Chong & Treisman, 2005). The present studies endeavored to directly investigate whether distributed, or focused attention was more easily combined with scene categorization. In Experiment 1, we demonstrated that judgments relating to mean circle size were faster when combined with a distributed attention task compared to a focused attention task. In Experiment 2, we extended this finding by showing that although the addition of a simultaneous attention task slowed responses, the ability to detect an animal was more easily combined with tasks requiring distributed rather than focused attention. In Experiments 3 and 4, we investigated whether this distributed advantage would extend to more complex natural scenes. When observers were asked to classify natural scenes based on their basic level, there was no statistically significant difference between distributed and focused attention task conditions (Experiment 3). However, when asked to categorize scenes based on the superordinate level, responses were faster when combined with a distributed attention task compared to a focused attention task (Experiment 4).

An interesting question is how exactly the extraction of a scene's summary statistics contributes to rapid scene categorization? One possibility that is consistent with the present results is that a scene's summary statistics contributes to the formation of a low-resolution global scene structure that allows for both superordinate and basic level categorization. It is argued that a scene's summary statistics contributes to the creation of a set of orthogonal global features that defines a scene's overall shape, such as its "degree of openness", "mean depth", and "navigability" (see e.g., Spatial Envelope Theory; Oliva & Torralba, 2001). Greene and Oliva (2009) suggested that whereas basic level categorization requires the integration of several global features (e.g., a forest is both "natural" and "closed"), superordinate level categorization (e.g., natural/manmade distinction) can be based on a single global feature (e.g., "naturalness"). Consistent with this suggestion, the "naturalness" global property hypothesized to differentiate between manmade and natural scenes correlates with low-level features that are distributed homogeneously over an image (Torralba & Oliva, 2003). Combined with the finding that superordinate level categorization occurs before basic level categorization (Loschky & Larson, 2010; Joubert et al., 2007; Rousselet, Joubert, & Fabre-Thorpe, 2005; Greeen & Oliva, 2009), these findings suggest that superordinate level information is available earlier than basic level information because there is no need to integrate global features. The present results provide further support for this hypothesis by demonstrating that distributed attention facilitates superordinate categorization of both natural scenes and objects, but not basic level categorization of natural scenes.

A related question to the one above is why basic level categorization is not facilitated by distributed attention? As previously mentioned, our preferred explanation is that there is an unequal amount of global features needed to classify superordinate and basic level categories (Greene & Oliva, 2009; Gosselin & Schyns, 2001). The potential for interference caused by conflicting global features is greater in basic level categorization compared to superordinate level categorization, because of the need to combine different source of information. Consistent with this idea, Greene and Oliva (2009) demonstrated that when observers are asked to categorize scenes at the basic level, the false alarm rate is greater when a distractor image shares global properties with the target category than when a distractor image does not (e.g., for the target category, forest, a distractor image that contained close space was more likely to produce a false alarm than a open space distractor image). This finding is consistent with studies investigating the perceptual averaging phenomenon, which is hypothesized to underpin the construction of global properties. Specifically, the presence of an irrelevant set of items influences mean judgments related to a relevant set, suggesting that the computation of the irrelevant set's mean interferes with response behaviour (Oriet & Brand, 2013). Similarly, observers are able to compute summary statistics of two sets of interspersed objects concurrently, but doing so incurs a cost of dividing attention across the two sets (Brand, Oriet, & Sykes-Tottenham, 2012; Emmanouil & Treisman, 2008). These results thus provide converging evidence that several summary descriptors of a scene are computed independently, but nevertheless interact causing interference effects to emerge during certain scene related judgments. Thus, it is possible that basic level categorization does not benefit from distributed attention because interference effects resulting from the integration of global features neutralize any benefits that result from adopting a distributed attention strategy.

An alternative explanation for the differing effects of distributed attention on basic and superordinate level categorization could be that there were different task demands between the experiments. As previously mentioned, there were more possible target scenes in our basic level scene categorization task, compared to our animal detection task and our superordinate level categorization task. Thus, we cannot conclude for certain that distributed attention facilitates only superordinate categorization. Such a claim would require the replication of Experiment 3 while controlling for the number of possible target categories. Nevertheless, we see the present results as converging with previous research in suggesting that both animal detection and superordinate level categorization occur earlier than basic level categorization because of their evolutionary importance (Loschky & Larson, 2010; Fei Fei et al., 2007). Fei Fei et al. (2007) reported that the earliest categorical distinction made between scenes is the superordinate indoor/outdoor distinction. Loschky and Larson (2010) subsequently suggested that the availability of superordinate level information is arranged into a hierarchy in which more primate distinctions (e.g., indoor/outdoor) are made prior to less primitive distinctions (manmade/natural). It would thus be interesting to determine whether distributed attention facilitates the indoor/outdoor distinction more than the natural/manmade distinction. Along the same lines, the superordinate animal/no animal distinction is argued to have significant evolution priority (Li et al., 2002). A similar interesting question would be to determine whether distributed attention influences evolutionary important object categories in the same fashion as non-evolutionary important categories (e.g., vehicle/no vehicle distinction).

A question the present studies were unable to address was why distributed attention facilitated superordinate scene categorization on both target present and target absent displays, but only animal detection on target present displays. Given that both tasks required superordinate level categorization, there should have been no difference between the tasks. However, this finding is consistent with emerging research showing that animal detection performance differs from scene categorization performance within the same task. Cohen and coworkers (2011) found that when observers were asked to complete an attentionally demanding task concurrently with a scene categorization task, the ability to detect the presence of animal was more susceptible to the costs of dividing attention than basic level scene categorization. Nevertheless, the animal detection results in Experiment 2 are consistent with our previous suggestion that the ability to detect an animal is an important evolutionary adaption. For example, it is more important to detect the presence of animal (target present displays) than to detect the absence of an animal (target absent displays), as indexed by the fact that there was no significant difference between distributed and focused attention tasks on target absent trials. A critical next step in advancing our understanding of scene categorization behaviour will be to investigate the distinction between rapid object categorization and more complex scene categorization.

Despite its rapid nature, manipulations of attention affect scene categorization performance. Thus, a challenge for scene perception researchers is to establish a theory of scene perception that includes attention, while also acknowledging its rapid nature. The present study provides a first step in addressing this issue by motivating observers to use focused, or distributed attention strategies and measuring the subsequent effects on scene categorization behaviour. Specifically, the present work examined whether the effects of distributed attention on scene categorization performance were the same as the effects of distributed attention on perceptual averaging, a mechanism hypothesized to contribute to rapid scene perception. As suggested by the present results, further research investigating the effects of attentional distribution on scene categorization performance could enrich our understanding of both mechanisms involved and the type of information that allows for rapid responses to scene stimuli.

4. Conclusions and Summary

A scene's semantic information facilitates target localization (Eckstein, Drescher, & Shimozaki, 2006; Torralba, Oliva, Castelhano, & Henderson, 2006), influences object recognition (Boyce & Pollatsek, 1992), and speeds up recall for previously memorized scenes (Brewer & Treyens, 1981; Pezdek, Whetstone, Reynolds, Askari, & Dougherty, 1989). On the basis of such evidence, it has been argued that the rapid extraction of a scene's meaning is the earliest meaningful stage of scene perception (Oliva, 2005). Accordingly, a large majority of researchers have focused on understanding scene perception by identifying the sources of information that underpin rapid scene categorization. Information related to an image's spatial frequencies (Oliva & Schyns, 1997; Schyns & Oliva, 1994), color (Castelhano & Henderson, 2008; 2005; Loschky & Simons, 2004; Oliva & Schyns, 2000), phase (Joubert, Rousselet, Fabre-Thorpe, & Fize, 2009; Loschky et al., 2007, 2010; Loschky & Larson, 2008), summary statistics (Evans & Treisman, 2005), and central and peripheral regions (Larson & Loschky, 2009) have all been suggested to contribute to the ability to rapidly categorize a scene. The purpose of the present dissertation was to build upon these findings by addressing how the information that is required to perform scene categorization is selected.

This question was investigated by evaluating the hypothesis that one of the mechanisms responsible for selecting scene information is attention (Oliva & Schyns, 1997; Treisman, 2006). Consequently, the experiments reported in this work departed from the traditional methodologies used to study the role of attention in scene categorization. Previous researchers relied on the use of dual task paradigms to investigate whether attention to scene stimuli is needed to extract semantic scene related information (as reviewed in the General Introduction of Chapter 1). Whereas some of these studies documented a cost of dividing attention (Cohen et al., 2011; Walker et al., 2008), others did not (Li et al., 2002; 2005). These two sets of conflicting findings have been particularly influential because they suggest that attention affects the ability to detect a scene, but only under specific circumstances (e.g., only under high attentional load). The present work sought to investigate how adopting different attentional orientations affected scene categorization. Specifically, it tested the hypothesis that one role of attention is to select scene information that is used to make rapid scene-related decisions. The experiments reported in this work tested this hypothesis by using novel experimental paradigms that combined experimental tasks from different cognitive domains. Whereas the paradigm in Chapter 2 combined hierarchical figure perception with scene categorization tasks, the paradigm in Chapter 3 combined a perceptual averaging task with scene categorization tasks. Although both chapters were based on different theories and relied on different methodologies, they nevertheless converged in suggesting that scene perception requires attention, and that one of its roles is to facilitate the extraction of information that is used to rapidly categorize scenes. A further benefit of Chapter 3 is that its design allowed for a discussion relating to the hypothesis that the information conveyed in overall summary statistics contributes to the formation of information in an image's spatial scales (e.g., The Spatial Envelope Theory discussed in section 1.4.3 of the General Introduction in Chapter 1; see also, Oliva & Torralba, 2001). In the following sections, I elaborate on the results of each of the experiments reported in Chapters 2 and 3, with particular emphasis on the significance of their findings and potential limitations that constrain their conclusions.

4.1 Review and significance of main findings

4.1.1 Chapter 2. Chapter 2 investigated Oliva and Schyn's (1997) hypothesis that attending locally and globally facilitates categorization based on an image's fine and coarse information, respectively. This hypothesis was investigated by priming observers to perceive scenes in either a global, or local fashion, by asking them to complete local and global Navon tasks (Navon, 1972). Overall, the results not only converged with pervious research in demonstrating that scene perception requires attention (Walker et al., 2008; Cohen et al., 2011; Evans & Treisman, 2005), but also suggested a novel role for attention. Specifically, that attention facilitates the selection of scene information. This conclusion was based on the collective findings of four experiments that measured spatial scale selection using hybrid images (i.e., a low-pass filtered image of one scene combined with a high-pass filtered image of a different scene). Although researchers showed that observers were able use both fine (conveyed by an images high spatial frequencies [HSFs]) and coarse (conveyed by an image's low spatial frequencies [LSFs]) information as the basis for hybrid image categorization (Schyns & Oliva, 1994; Oliva & Schyns, 1997), it was nevertheless unknown whether there was a preferred spatial scale. Low spatial frequencies were argued to form the basis of scene categorization because they are available earlier in the visual system than HSFs (Schyns & Oliva, 1994; Morrison & Schyns, 2001). However, HSFs have been shown to be the preferred spatial scale under certain task constraints (Oliva & Schyns, 1997); suggesting that differences in LSF- and HSF-based hybrid categorization is a function of task demands. Thus, Experiments 1 and 2 were designed to address whether LSFs or HSFs are the preferred spatial scale, irrespective of any experimental manipulation. Experiment 1 assessed spatial scale sensitivity by asking observers to indicate whether a cue word corresponded to the category of the previously presented image. The results of Experiment 1 replicated the finding that observers are sensitive to both LSF and HSF information (Oliva & Schyns, 1997; Schyns & Oliva, 1994). Experiment 2 extended this result by investigating spatial scale preference by asking observers to identify the category of the previously presented scene by choosing from a list of all possible target categories. The results of Experiment 2 demonstrated that observers overwhelmingly selected the target category that corresponded to LSF information, despite the fact that observers were more sensitive to a hybrid's HSFs (Experiment 1). These results corroborate previous research that suggested a critical role for LSFs in rapid scene categorization (Loschky & Simons, 2004; McCotter, Gosselin, Sowden, & Schyns, 2005; Schyns & Oliva, 1994; Oliva & Schyns, 1997), and extended these findings by demonstrating a preference to use LSF information despite the fact that HSF information is more salient.

The results of Experiments 1 and 2 provided a baseline pattern of performance that could be used to answer the central question of chapter 2: does attention to local and global levels of a scene bias the selection of spatial scale? This question was addressed in Experiment 3 by asking observers to complete either global or local Navon tasks prior to categorizing hybrid images. Because observers preferred to categorize hybrids based on LSF content, it was hypothesized that LSF-based hybrid categorization would be facilitated following global, but not local Navon tasks. Consistent with this hypothesis, observers preferred to base hybrid categorization on LSF content, and LSF-based hybrid categorization was faster following global Navon tasks compared to local Navon tasks. Thus, the contributions of Experiment 3 to the scene categorization literature have been two-fold. First, the results provide empirical evidence that can help elucidate the debate regarding whether attention is needed in order to consciously perceive a scene. Specifically, the results corroborate previous studies in suggesting that scene categorization can benefit from attention (Cohen et al., 2011; Evans & Treisman, 2005; Walker et al., 2008). Second, the results of Experiment 3 suggest a novel role for attention; particularly, that one role of attention is to facilitate the selection of scene information. This finding dovetails with Larson and colleagues (2014), who showed that manipulations of selective attention affect the selection of central and peripheral scene information.

A limitation to the conclusions of Experiment 3 is that the analogy between local and global processing and fine and coarse processing is confounded by the fact that both are associated with HSFs, and LSFs, respectively. Thus, it is unclear whether the observed LSF-based hybrid categorization facilitation effect following global processing was due to the priming of a hybrid's LSFs, or the priming of a hybrid's coarse information. To address this issue, Experiment 4 replicated Experiment 3; however LSFs were suppressed in the Navon figures by contrast balancing the Navon stimuli. Similar to Experiment 3, observers preferred to base hybrid categorization on LSF information; however and in contrast, LSF-based hybrid categorization was faster following local Navon tasks. One interpretation of this result is that the removal of a Navon's LSFs primed the selection of different information within the LSF content of scenes. This interpretation is consistent with Oliva and Schyns (1997), who claimed that fine and coarse information is orthogonal to local and global information; that is, there is more than one type of information at both local and global scales and it is possible to direct attention to these differing sources of information. Thus, Experiment 4 not only corroborates the finding that observers prefer to use a single spatial scale as the basis for scene categorization (Experiments 2 and 3; Oliva & Schyns, 1997), but it is also the first set of results to empirically support Oliva and Schyns' claim that there is more than one source of information at each spatial scale.

Although Chapter 2 provides a significant contribution to the scene perception literature, it nevertheless has a number of limitations. One limitation is that the spread of attention potentially differed between local and global conditions in Experiments 2 and 3. It is possible that attending to the global level of a Navon figure results in a wider spread of attention compared to the more narrowed spread of attention that results from attending to the local level. A consequence of this possibility is that stimulus size, and not attention to hierarchical level, is the critical dimension affecting scene categorization performance. However, this possibility is unlikely for two main reasons. First, Navon stimuli were purposely presented in the periphery and observers were instructed to make same/different judgments, prompting a scan across the entire visual field. Second, Flevaris and colleagues (2011) ruled out this possibility in their investigation on how attention to hierarchical level affected spatial scale selection in sine wave gratings (see the Introduction to Chapter 2). Flevaris and colleagues conducted a control experiment in which they replaced the Navon letters with single letters that were the size of the global Navon letters, or the size of the local Navon letters. There was no statistically significant difference on observers' ability to detect both low, and HSFs between local- and global-sized letter conditions, suggesting that attention to hierarchical level was required to facilitate the selection of LSFs and HSFs, respectively.

A second limitation of the experiments in Chapter 2 is that natural images have on average more power (i.e., more energy) in the low spatial frequencies than in the high spatial frequencies (Hansen, Haun, & Essock, 2008). This dichotomy was not controlled for in these studies. Thus, it is possible that observers preferred to categorize hybrids based on LSFs because they contained more power than HSFs. However, given that this LSF/HSF dichotomy is a natural phenomenon, I was hesitant to alter the scene stimuli for fear of manipulating the appearance of the scenes. Manipulations of the amplitude spectra slope in artificial images does not appear to affect the perception of images; however, in natural images, Johnson, Richard, Hansen, and Ellemberg (2011) showed that artificial manipulations of the amplitude spectra slope results in the images looking unnatural. Specifically, changes that cause an overrepresentation of the LSF contrast energy (i.e., a steeper amplitude spectrum slope) cause images to be perceived as more blurred. Conversely, changes that cause an increased representation of the HSF contrast energy (i.e., a shallower amplitude spectrum slope) result in a "whitened" image perception.

A third limitation of Chapter 2 (and Chapter 3) is that the use of white noise masks (amplitude spectrum slope = 0; maximum orientation = 0) produced weak backward masking effects. A potential consequence of weak backward masking effects is that the effects of attention on scene categorization were not strictly related to early visual processes, which have been hypothesized to contribute to rapid scene categorization. A white noise mask was chosen in order to obtain an approximate 80% scene categorization accuracy rate, as reported by Hansen and Loschky (2013). However, accuracy in the present set of studies exceeded this level. A review of the literature suggests that a more effective way to determine a specific level of performance would be to pilot test the target/mask duration ratio in order to establish a wide range of useful performance measures. This technique has been successfully employed in previous research examining scene categorization performance (see e.g., Figure 10 in Loschky et al., 2007). Furthermore, the present set of studies varied the presentation duration of scene images while holding the mask duration constant. Breitmeyer and Ogmen (2000) suggested that a more effective masking paradigm is to hold both the target image and the mask duration constant and vary the interstimulus interval between target and mask.

***4.1.2 Chapter 3*.** Similar to Chapter 2, Chapter 3 addressed the hypothesis that attention facilitates the selection of scene information. However, whereas Chapter 2 addressed this question with respect to the selection of spatial scale, Chapter 3 addressed this question with respect to the extraction of overall summary statistics. Specifically, it was designed to address Evans and Treisman's (2005) claim that distributed attention facilitates that rapid extraction of summary statistics that underpins rapid scene categorization. Four experiments investigated whether manipulations of attention known to affect the extraction of summary statistics also affected rapid scene categorization in the same fashion. The results were overall consistent with Chapter 2 in suggesting that one role of attention is to facilitate the selection of information that is used to make rapid scene categorizations. Although it is suggested that summary statistics provide information used in rapid scene categorization, the majority of this evidence is based on indirect observations. For example, the two phenomena have been linked because both are computed rapidly and appear to provide overall global properties that are extracted without the need for attention (Evans & Treisman, 2005; Greene & Oliva, 2009). The lack of direct empirical evidence is due, in part, to the fact that the paradigms used to study each respective phenomenon are not easily comparable. For example, they often use different dependent variables. Thus, the purpose of Experiment 1 was to two-fold: 1) to replicate Chong and Treisman's (2005) claim that adopting a distributed attention mode facilitates the extraction of statistical properties from a set of similar items; and 2) to modify their original design so that the results could be more easily compared to scene categorization studies. In pursuit of these goals, Experiment 1 used a modified version of Chong and Treisman's original task in order to focus on RT measures rather than threshold differences. Consistent with Chong and Treisman's original report, observers were faster at judging which of two test dots corresponded to the mean size of a preceding set of dots when combined with a task requiring distributed attention compared to a task requiring focused attention. Thus, the results of Experiment 1 not only corroborate Chong and Treisman's original result, but also establish a baseline pattern of results to which the effects of attention on scene categorization could be compared. The logic behind this comparison is that if distributed attention facilitates the extraction of summary statistics, which in turn, facilitates rapid scene categorization, then the effects of distributed attention on scene categorization should be the same as the effects of distributed attention on summary statistics. Experiment 2 was designed to test this hypothesis by replicating Experiment 1 with the exception that the perceptual averaging task was replaced with an animal detection task, which had been previously used as a measure of scene categorization behaviour (e.g., Li et al., 2002; 2005; Rousselet et al., 2002). Consistent with Experiment 1, the ability to detect the presence of an animal was faster when combined with a distributed attention task, in comparison to a focused attention task. To my knowledge, this is the first empirical demonstration examining summary statistics in conjunction with scene categorization.

Early investigations into the effects of attention on scene categorization have focused on the ability to detect the presence of an animal. However, as scene perception studies have evolved, researchers have used more complex scene stimuli. This is done to investigate if the results obtained using simple animal stimuli (e.g., Fei Fei et al., 2002; 2005) would extend to more complex natural scenes. Thus, the purpose of Experiments 3 and 4 was to investigate whether distributed attention facilitates natural scene categorization in the same fashion as Experiments 1 and 2. Observers in Experiments 3 and 4 were asked to complete concurrent distributed and focused attention tasks with scene categorization tasks that required either basic (Experiment 3) or superordinate (Experiment 4) level categorization. When observers categorized images based on their basic level (e.g., beach and forest), there was no statistically significant difference between the effects of focused and distributed attention tasks; however, superordinate level categorization was faster when combined with a distributed attention task than a focused attention task. Together, the results of Experiments 3 and 4 suggest that attention facilitates the selection of information used for superordinate, but not basic level categorization. This finding corroborates previous research that suggests superordinate categorization occurs before basic level categorization (Joubert et al., 2007; Rousselet, Joubert, & Fabre-Thorpe, 2005; Larson & Loschky, 2009), and extends these findings by suggesting that attention is more likely to influence the selection of superordinate level information since it is available earlier than basic level information.

The design of Chapter 3 also allowed for a discussion about the association between summary statistics and the information contained in spatial scales, as discussed in Chapter 2. Similar to Evans and Treisman's (2005) suggestion that summary statistics contributes to the formation of disjunctive features, Oliva and Torralba (2006) and Greene and Oliva (2009) suggested that a scene's summary statistics contributes to the formation of global features that are carried in an image's spatial scales. Evidence in favour of this hypothesis has been limited to simulation studies (e.g., Spatial Envelope Model; Oliva and Torralba, 2001) and to correlational evidence that has shown that the Fourier spectra associated with various global properties correlates with low-level features that are distributed homogeneously over an image (Torralba & Oliva, 2003). The experiments described in Chapter 3 provide behavioural evidence for this hypothesis by providing data that supports certain predictions based on the Spatial Envelope Model. For example, the Spatial Envelope Model predicts that a scene's summary statistics contributes to the formation of global features, which suggests that distributed attention should not only facilitate perceptual averaging, but also scene categorization. Chapter 3 provides direct evidence for this prediction by demonstrating that the effects of completing a concurrent distributed attention task are the same between scene categorization tasks and perceptual averaging tasks (at least for superordinate level categorization).

Another prediction of the Spatial Envelope Model is that a single global feature is sufficient for superordinate, but not basic level categorization. Consistent with this prediction, several researchers have shown that superordinate level information is available before basic level information (Joubert et al., 2007; Rousselet, Joubert, & Fabre-Thorpe, 2005; Larson & Loschky, 2009), suggesting that only superordinate level information is available pre-attentively. If distributed attention facilitates the processing of pre-attentive information as hypothesized by Evans and Treisman (2005), then distributed attention should only facilitate superordinate categorization. This logic is based on the premise that the time required to integrate global features associated with basic level categorization would extend pass the time thought to be pre-attentive. In support of this hypothesis, the results of Chapter 3 Experiments 2 and 3 demonstrated that distributed attention facilitates superordinate level categorization of both animals and natural scenes, but not basic level categorization of natural scenes.

Similar to Chapter 2, there are several limitations that must be considered when evaluating the conclusions of Chapter 3. One limitation is that the difficulty of the attention task was not standardized across participants. This is problematic because the difficulty of a secondary attention task is a critical factor affecting scene categorization performance (Cohen et al., 2011; Walker et al., 2008). This is particularly concerning in Experiment 4 because performance on the distributed attention task was statistically significantly higher than performance on the focused attention task. However, there are two main reasons why differences in attention task difficulty were unlikely to have significantly affected scene categorization performance in Chapter 3. First, attention task accuracy was high in all experiments and there were no significant differences between distributed and focused attention tasks in Experiments 1 - 3, suggesting that both tasks were equally difficult. Second, although distributed attention task accuracy was statistically significantly higher than focused attention task accuracy in Experiment 4, accuracy was greater than 90% for both tasks. The fact that accuracy was high for both attention tasks suggests that any differences between the two had limited effects on scene categorization performance.

In order to control for attention task difficulty, Chong and Treisman (2005) titrated the ratio of both the small and large rectangles in order to maintain a standard level of correct performance. This ensured that any observed difference on the mean discrimination task between distributed and focused attention conditions was not a function of attention task difficulty. The primary reason why a similar staircase method was not used in the present studies is that changing the aspect ratio of the rectangles would potentially, and unnecessarily, direct attentional resources to the attention task. This is worrisome because the addition of secondary attention task affects scene categorization performance, with larger effects associated with increased task difficulty (Cohen et al., 2011; Walker et al., 2008). As such, I wanted to ensure that both attention tasks were relatively easy, without drawing attention to the attention tasks. Early pilot testing indicated that attention task performance was greater than 85% for both focused and distributed attention tasks, so the decision was made not to titrate the aspect ratios of the rectangles. Nevertheless, attention task accuracy varied between the experiments reported in Chapter 3. Therefore, it is likely that some of the observed scene categorization differences found between the experiments can be attributed to differences in attention task difficulty. Similar to the masking problem in Chapter 2, a solution to this problem would be to pilot test a combination of rectangle ratios for each scene categorization task. This would not only ensure that attention task accuracy would be equal between distributed and focused conditions, but also that the association between attention task difficulty and scene categorization difficulty would be consistent between experiments. Similarly, observers could undergo baseline testing prior to the experiment in order to determine their individual thresholds for the distributed and focused attention tasks (e.g., determine the aspect ratio for each individual that corresponds to 75% accuracy). However, research being conducted in Concordia's vision laboratory has shown that this procedure is time consuming, and that individual thresholds vary from time to time (and day to day).

A second limitation of Chapter 3 is that perceptual averaging RTs in Experiment 1 were overall much slower than scene categorization RTs in Experiments 2 - 4. This finding is most likely caused by methodological differences. Whereas Experiment 1 required observers to make a comparison between two test probes, Experiments 2 – 4 required the evaluation of a single cue word. Thus, the extra time needed to evaluate a second probe item may have caused the increased RTs observed in Experiment 1. Future research that investigates similarities between perceptual averaging and scene categorization paradigms should take such task demands into consideration. With respect to the present study, a more comparable perceptual averaging paradigm would have been to index mean size extraction using a single test probe. In his seminal investigation, Ariely (2001) used a single probe design in which he asked his observers to indicate whether a test dot was smaller, or larger than the average size of a preceding set of dots. Despite this limitation, the primary objective of the Chapter 3 was to demonstrate that distributed attention affected perceptual averaging in the same fashion as rapid scene categorization. As such, the most relevant comparison is the pattern of results between experiments, and not absolute time.

4.3 Future directions.

The experiments reported in this dissertation corroborate previous research that scene categorization can benefit from attention (Cohen et al., 2011; Walker et al., 2008;

Evans & Treisman, 2005) and extend these finding by providing evidence that one role of attention is to select scene information. This investigation examined the effects of attention on the selection of a scene's spatial scales and overall summary statistics; however there are other additional features hypothesized to contribute to scene categorization. Thus, future research should investigate how attention affects the selection of other features not investigated here. Recent evidence suggests that such investigations can be beneficial in elucidating attention's role in scene categorization. For example, Larson and colleagues (2014) found that manipulations of selective attention affect the selection of central and peripheral scene information. The authors subsequently argued that understating how attention affects the processing and selection of such scene information is crucial in developing computational models of scene perception that can lead to a fuller understanding of the efficiency with which humans accomplish rapid scene categorization.

One particular area of interest is the selection of scene colours. Studies that have investigated scene categorization (including the present body of work) have typically used gray-scaled images in order to control for the mediating effects of colour on scene recognition (Oliva & Schyns, 2000). However, this technique is shortsighted given that colour provides information that can be used to categorize natural scenes (e.g., Castelhano & Henderson, 2008; Goffaux et al., 2005; Loschky & Simons, 2004; Oliva & Schyns, 2000). The finding most relevant to the present discussion is the suggestion that colour information contributes to the formation of coarse information that Chapter 2 demonstrated to be the preferred spatial scale. Sanocki and Epstein (1997) suggested that a scene's global layout is due, in part, to the organization of coloured blobs within an image's coarse scale. For example, an image of a beach can be identified based on the fact that the sky and the ground can be differentiated on coloured blobs. On the basis of this evidence, Oliva and Schyns (2000) suggested that colour information rapidly facilitates the segmentation of an image, which in turn, activates a scene schema that can be used as the basis for categorization. Combined with the fact that endogenous attention affects the perception of saturation (Fuller & Carrasco, 2006), and there is good evidence that attention could modulate how colour information is used in scene categorization.

Another source of information argued to be important for scene categorization is an image's phase (i.e., the distribution of spatial frequencies and orientations in an image with respect to their location; Joubert, Rousselet, Fabre-Thorpe, & Fize, 2009; Loschky et al., 2007, 2010; Loschky & Larson, 2008). However, to date, only a single study investigated the effects of attention on the processing of phase information (Kihara & Takeda, 2012). Similar to the majority of research looking at the effects of attention on scene categorization, Kihara and Takeda (2012) showed that phase information is processed pre-attentively; nevertheless, and as demonstrated by the present set of studies, this finding does not preclude the possibility that attention can facilitate the selection of phase information.

One of the most interesting findings from the present work is the differing effect of attending globally on LSF-based hybrid categorization seen between Experiments 3 and 4 in Chapter 2. Whereas LSF-based hybrid categorization was faster following global Navon tasks in Experiment 3, it was slower following global Navon tasks in Experiment 4. As previously mentioned, this finding can be interpreted to suggest that the processing of a Navon's LSFs in Experiment 3 had a significant impact on the selection of spatial scale. The present studies were not designed to investigate this issue. One technique that has been successfully shown to elucidate the processing of spatial frequencies in hierarchical figures is event-related brain potentials (ERPs). Flevaris, Martinez, and Hillyard (2014) reported that ERP activity elicited by spatial frequency gratings differed as a function of attended Navon level. Specifically, attending to global levels facilitated the processing of LSF gratings with ERPs differing between global and local Navon tasks at approximately 196–236 msec after stimuli onset. Furthermore, this difference was concentrated over the right occipital scalp. In contrast, attending to local levels of Navon stimuli facilitated the processing of HSF gratings with differences in ERPs between global and local Navon tasks occurring at approximately 250–290 msec after stimulus onset. In contrast to the LSF gratings, these differences were distributed over the entire occipital scalp. Thus, a possible way to elucidate how the processing of a Navon's LSFs affected the subsequent categorization of hybrid images would be to investigate the differential processing of ERPs elicited by contrast balanced Navon stimuli and full broadband Navon stimuli.

4.4 Concluding remarks

Hierarchical structure is a common occurrence in our visual environment. At the top of the hierarchy is a global structure (e.g., a forest), which is comprised of a local structure (e.g., tress), which in turn, is comprised of an even more local structure (e.g., leaves), and so on. As we navigate through our environment, we are constantly switching our focus of attention extracting information from each attended level with effortless ease. At any given point, we are able to focus our attention on a particular object, such as a tree, or adopt a global spread of attention to see the forest. An important question for psychologists has been to determine what level of information is available during the very first glimpse of a scene. Over the past 40 years, findings from psychophysics, psychophysiological, and simulation studies have converged on the conclusion that the first meaningful stage of natural scene perception is the extraction of a scene's global meaning (e.g., you do see the forest before the trees). At the same time, these studies have suggested that the information that underpins the ability to rapidly categorize a scene originates from early visual processes. The present dissertation was a first attempt to investigate how attending locally and globally to a scene affected the selection of low-level visual scene information. Although the results did not provide conclusive evidence that attention is needed in order to extract global semantic content, they nevertheless demonstrated that one of the primary roles of attention is to facilitate the selection of scene information that can be used to extract semantic meaning.

References

Alvarez, G. A., & Oliva, A. (2008). The representation of simple ensemble features outside the focus of attention. *Psychological Science, 19*, 392 - 398

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science, 12,* 157–162. doi:10.1111/1467-9280.00327.

Badcock, J., Whitworth, F., Badcock, D., & Lovegrove, W. (1990). Low-frequency filtering and the processing of local-global stimuli. *Perception*, *19*, 617–629.

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

Block, N. (1995). On a confusion about a function of consciousness. *Behavioral & Brain Sciences*, *18*, 227–247.

Boyce, S. J., & Pollatsek, A. (1992). Identification of objects in scenes: The role of scene background in object naming. *Journal of Experimental Psychology: Learning Memory, and Cognition, 18,* 531–543. doi: 10.1037/0278-7393.18.3.531.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436.

Brand, J., Johnson, A., & Von Grunau, M. (2012). Global, but not focused attention impairs the ultra-rapid categorization of natural scenes [Abstract]. *Journal of Vision*, *12*, 796. doi: 10.1167/12.9.796.

Brand, J., Oriet, C., & Sykes-Tottenham, L. (2012). Size and emotion averaging: Costs of dividing attention after all. *Canadian Journal of Experimental Psychology, 66*, 63-69. doi:10.1037/a0026950.

Braun, J. (2003). Natural scenes upset the visual applecart. *Trends in Cognitive Science*, *7*, 7 – 9.

Brewer, W. F., & Treyens, J. C. (1981). Role of schemata in memory for places. *Cognitive Psychology, 13,* 207–230. doi:10.1016/0010- 0285(81)90008-6.

Brown, T. L., Gore, C. L., & Carr, T. H. (2002). Visual attention and word recognition in stroop color naming: Is word recognition “automatic?” *Journal of Experimental Psychology: General, 131,* 220–240. doi: 10.1037/0096 3445.131.2.220.

Cameron, E. L., Tai, J. C., & Carrasco, M. (2002). Covert attention affects the psychometric function of contrast sensitivity. *Vision Research, 42*, 949–967.

Campbell, W. F., & Robson, G. J. (1968). Application of Fourier analysis to the visibility of gratings. *Journal of Physiology*, *197*, 551-566.

Castelhano, M. S., & Henderson, M. J. (2008). The influence of color on the perception of scene gist. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 660- 675. doi: 10.1037/0096-1523.34.3.660.

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research*, *43*, 393–404.

Chong, S. C., & Treisman, A. (2005). Attentional spread in statistical processing of visual displays. *Perception & Psychophysics, 67,* 1–13. doi:10.3758/BF03195009.

Choo, H., & Franconeri, S. L. (2010). Objects with reduced visibility still contribute to size averaging. *Attention, Perception, and Psychophysics, 72,* 86–99. doi:10.3758/APP.72.1.86.

Cohen, A. M., Alvarez, A. G., & Nakayama, K. (2011). Natural-scene perception requires attention. *Psychological Science*, *22*, 1165-1172.

Corbett, J. E., & Oriet, C. (2011). The whole is indeed more than the sum of its parts: Perceptual averaging in the absence of individual item representation. *Acta Psychologica, 138,* 289 –301. doi:10.1016/j.actpsy.2011.08.002.

Corel. (1996). Corel stock photo library [CD-ROM]. Ottawa, Ontario, Canada: Author.

Crouzet, S., Kirchner, H., & Thorpe, S. J. (2010). Fast saccades toward faces: Face detection in just 100 ms. *Journal of Vision*, 10(4):16, 1-17. doi:10.1167/10.4.16.

de Fockert, J., & Wolfenstein, C. (2009). Rapid extraction of mean identity from sets of faces. *Quarterly Journal of Experimental Psychology*, *62*, 1716–1722. http://dx.doi.org/10.1080/17470210902811249.

de Gardelle, V., & Kouider, S. (2009). How spatial frequencies and visual awareness interact during face processing. *Psychological Science*, *21*, 58-66.

Eckstein, M. P., Drescher, B. A., & Shimozaki, S. S. (2006). Attentional cues in real scenes, saccadic targeting, and Bayesian priors. *Psychological Science, 17,* 973 980. doi:10.1111/j.1467-9280.2006.01815.

Emmanouil, T. A., & Treisman, A. (2008). Dividing attention across feature dimensions in statistical processing of perceptual groups. *Perception & Psychophysics, 70*, 946–954. http://dx.doi.org/10.3758/PP.70.6.946.

Evans K.K., & Chong S. (2011). *Distributed Attention and Its Implication For Visual Perception*. In Wolfe, J.M., Robertson, L. (Eds.). Attention and Beyond: Searching through the papers of Anne Treisman. NY: Oxford University Press.

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

Fabre-Thorpe M., Delorme A., Marlot C. & Thorpe S.J. (2001). A limit to the speed of processing in ultra-rapid visual categorisation 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):10, 1–29, http://journalofvision.org/7/1/10/, doi:10.1167/7.1.10.

Findlay, J. M., Brogan, D., & Wenban-smith, M. G. (1993). The spatial signal for saccadic eye movements emphasizes visual boundaries. *Perception and Psychophysics*, *53*, 633-641.

Flevaris, V. A., Bentin, S., & Robertson, C. L. (2011). Attention to hierarchical level Influences attentional selection of spatial scale. *Journal of Experimental Psychology: Human Perception and Performance*, *37*, 12-22.

Flevaris, A. V., Martinez, A., & Hillyard, A. S. (2014). Attending to global versus local stimulus features modulates neural processing of low versus high spatial frequencies: an analysis with event-related brain potentials. *Frontiers in Psychology: Perception Science.* Retrieved from: http://journal.frontiersin.org/Journal/10.3389/fpsyg.2014.00277/full.

Fuller, S., & Carrasco, M. (2006). Exogenous attention and color perception: Performance and appearance of saturation and hue. *Vision Research*, *46*, 4032 4047.

Ginsburg, A. P. (1986). *Spatial filtering and visual perception*. In L. K. J. T. K. Boff (Ed), Handbook of Perception and Human Performance (Vol. II, Chapter 34, pp. 1-41). NY: Wiley.

Goffaux, V., Hault, B., Michel, C., Vuong, Q. C., and Rossion, B. (2005). The respective role of low and high spatial frequencies in supporting configural and featural processing of faces. *Perception 34*, 77–86.

Greene, R. M., & Fei Fei, L. (2013). Visual categorization is automatic and obligatory: Evidence from stroop-like paradigms. *Journal of Vision*, 14(1):14, 1-11. doi:10.1167/14.1.14.

Greene, M. R., & Oliva, A. (2009). Recognition of natural scenes from global properties: Seeing the forest without representing the trees. *Cognitive Psychology*, *58*, 137–176.

Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology*, *17*, 751–753.

Haberman, J., & Whitney, D. (2009). Seeing the mean: Ensemble coding for sets of faces. *Journal of Experimental Psychology: Human Perception and Performance, 35,* 718–734. doi:10.1037/a0013899.

Haberman, J., & Whitney, D. (2010). The visual system discounts emotional deviants when extracting average expression. *Attention, Perception, & Psychophysics, 72*, 1825–1838.

Hansen, C. B., & Loschky, C. L. (2013). The contribution of amplitude and phase spectra-defined scene statistics to the masking of rapid scene categorization. Journal of Vision, 13(13):21. Retrieved from: http://www.journalofvision.org/content/13/13/21.full.pdf.

Intraub, H. (1999). *Understanding and remembering briefly glimpsed pictures: Implications for visual scanning and memory.* In Coltheart, V. (Eds). Fleeting Memories: Cognition of Brief Visual Stimuli. MIT Press/Massachusetts

Jacoby, O., Kamke, M., & Mattingley, J. B. (2012). Is the whole really more than the sum of its parts? Estimates of average size and orientation are susceptible to object substitution masking. *Journal of Experimental Psychology: Human Perception and Performance, 39,* 233-244. doi: 10.1037/a0028762.

Joubert, O. R., Rousselet, G. A., Fabre-Thorpe, M., & Fize, D. (2009). Rapid visual categorization of natural scene contexts with equalized amplitude spectrum and increasing phase noise. *Journal of Vision, 9*, 1–16.

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

Kihara, K., & Takeda, Y. (2012). Attention-free integration of spatial frequency-based information in natural scenes. *Vision Research, 65,* 38-44.

Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Research*, *46*, 1762–1776

Koch, C., & Tsuchiya, N. (2007). Attention and consciousness: Two distinct brain processes. *Trends in Cognitive Sciences*, *11*, 16–22.

Koch, K., McLean, J., Segev, R., Freed, A. F., Berry II, J. M., Balasubramanian, V., & Sterling, P. (2006). How much can the eye tells the brain. Current Biology, *16*, 1428-1434.

Lamb, M. R., & Yund, E. W. (1993). The role of spatial frequency in the processing of hierarchically organized stimuli. *Perception & Psychophysics*, *54*, 773-784

Lamme, V. A. (2003). Why visual attention and awareness are different. *Trends in Cognitive Sciences*, *10*, 204–211.

Larson, A. M., & Loschky, L. C. (2009). The contributions of central versus peripheral vision to scene gist recognition. *Journal of Vision, 9*, 1–16.

Larson, A. M., Freeman, T. E., Ringer, R. V., & Loschky, L. C. (2014). The spatiotemporal dynamics of scene gist recognition. *Journal of Experimental Psychology: Human Perception and Performance, 40,* 471-487. DOI: 10.1037/a0034986.

Legge, G.E. & Gu, Y. (1989). Stereopsis and contrast. *Vision Research*, *29*, 989-1004.

Li, Fei Fei., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the Natural Academy of Sciences of the United States of America*, 99, 9596-9601. doi: 10.1073/pnas.092277599.

Loftus, R. G., & Masson, E. J., M. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, *1*, 476-490.

Losada, M. A., & Mullen, K. T. (1995). Color and luminance spatial tuning estimated by noise masking in the absence of off-frequency looking. *Journal of the Optical Society of America A*, 12(2), 250–260.

Loschky, L. C., Hansen, B. C., Sethi, A., & Pydimari, T. (2010). The role of higher-order image statistics in masking scene gist recognition. *Attention, Perception, & Psychophysics, 72,* 427–444. doi:10.3758/APP .72.2.427.

Loschky, L. C., & Larson, A. M. (2008). Localized information is neces- sary for scene categorization, including the natural/man-made distinction. *Journal of Vision, 8*(1), 1–9.

Loschky, L. C., & Larson, A. M. (2010). The natural/ man-made distinction is made before basic-level distinctions in scene gist processing. *Visual Cognition*, *18*, 513–536.

Loschky, L. C., Sethi, A., Simons, D. J., Pydimari, T., Ochs, D., & Corbeille, J. (2007). The importance of information localization in scene gist recognition. *Journal of Experimental Psychology: Human Perception & Performance, 33*, 1431-1450.

Loschky, L. C., & Simons, D. J. (2004). The effects of spatial frequency content and color on scene gist perception [Abstract]. *Journal of Vision, 4*(8), 881. doi:10.1167/4.8.881.

Mack, A., & Clarke, J. (2012). Gist perception requires attention. *Visual Cognition*, *20*, 300-327.

Mack, A., & Rock, I. (1998). *Inattentional blindness: Perception without attention*. Cambridge, MA: MIT Press.

Marois, R., Yi, D.-J., & Chun, M. M. (2004). The neural fate of perceived and missed events in the attentional blink. *Neuron*, *41*, 464–472.

Marshall, A. J., Burbeck, A. C., Ariely, D., Rolland, P. J., & Maritn, E. K. (1996). Occlusion edge blur: a cue to relative visual depth. *Journal of the Optical Society of America*, *13*, 681-688.

Martin, D., Slessor, G., Allen, R., Phillips, L. H. & Darling, S. (2012). Processing orientation and emotion recognition. *Emotion*, *12*, 39-43.

Morgan, M. J. (1992). Spatial filtering precedes motion detection. *Nature*, *355*, 344–346.

Morrison, J. D., & Schyns, G. P. (2001). Usage of spatial scales for the categorization of faces, objects, and scenes. *Psychonomic Bulletin & Review*, *8*, 454-469.

Mullen, T. K., & Losada, A. M. (1999). The spatial tuning of color and luminance peripheral vision measured with notch filtered noise masking. *Vision Research*, *39*, 721-731.

Myczek, K., & Simons, D. J. (2008). Better than average: Alternatives to statistical summary representations for rapid judgments of average size. *Perception & Psychophysics, 70,* 772–788. doi:10.3758/PP.70.5.772.

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

Oliva, A. (2005). Gist of the scene. in the *Encyclopedia of Neurobiology of Attention*. L. Itti, G. Rees, and J.K. Tsotsos (Eds.), Elsevier, San Diego, CA (pages 251-256).

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

Oliva, A., & Schyns, P. G. (2000). Diagnostic colors mediate scene recognition. *Cognitive Psychology, 41,* 176 –210. doi:10.1006/cogp .1999.0728.

Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, *42*, 145–175.

Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. *Progress in Brain Research: Visual Perception*, *155*, 23–36.

O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral & Brain Sciences*, *24*, 939–973

Oriet, C., & Brand, J. (2013). Size averaging of irrelevant stimuli cannot be prevented. *Vision Research*, *79*, 8-16.

Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience*, *4*, 739 744.

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. doi:10.1037/0278-7393.15.4.587.

Posner, M. I. (1994). Attention: The mechanisms of consciousness. *Proceedings of the National Academy of Sciences, USA*, *91*, 7398–7403.

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

Potter, M. C. (1999). Understanding sentences and scenes: The role of Conceptual Short Term Memory. In V. Coltheart (Ed.), *Fleeting memories: Cognition of brief visual stimuli* (pp.13-46). Cambridge, MA: MIT Press.

Potter, M.C., & Levy, E.I. (1969). Recognition memory for a rapid sequence of pictures. *Journal of Experimental Psychology, 81*, 10-15.

Raymond, E. J., Shapiro, L., & Arnell, M. K. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology: Human Perception and Performance, 18*, 849-860.

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.

Robitaille, N., & Harris, I. M. (2011). When more is less: Extraction of summary statistics benefits from larger sets. *Journal of Vision, 11,* 1–8. doi:10.1167/11.12.18.

Rousselet G., Fabre-Thorpe M., & Thorpe S.J. (2002). Parallel processing in high level categorization of natural images. *Nature Neuroscience*, *5*, 629-630

Sanocki, T., & Epstein, W. (1997). Priming spatial layout of scenes. *Psychological Science, 8*, 374–378.

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

Schyns, P. G., & Oliva, A. (1999). Dr. Angry and Mr. Smile: When categorization flexibly modifies the perception of faces in rapid visual presentations. *Cognition*, *69*, 243-265.

Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proceedings of the Natural Academy of Sciences of the United States of America*, *104*, 6424–6429.

Shulman, G. L., & Wilson, J. (1987). Spatial frequency and selective attention to local and global information. *Perception*, *16*, 89-101.

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

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

Slagter, H. A., Johnstone, T., Beets, I. A. M., & Davidson, R. J. (2010). Neural competition for conscious representation across time: An fMRI study. *PLoS ONE*, *5*(5), e10556. doi:10.1371/journal.pone.0010556.

Stroop, J. R. (1935). Studies of interfernce in serial verbal reactions. *Journal of Experimental Psychology*, *18*, 643-662. doi: 10.1037/h0054651.

Stromeyer, C. F., & Julesz, B. (1972). Spatial-frequency masking in vision: Critical bands and spread of masking. *Journal of the Optical Society of America A*, *62*, 1221–1232.

Thorpe, S. J., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*, 520–522.

Tononi, G., & Koch, C. (2008). The neural correlates of consciousness: an update. *Annals of The New York Academy of Sciences*, *1124*, 239-261.

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

Treisman, A. (2006). How the deployment of attention determines what we see. *Visual Cognition*, *14*, 411-443.

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

VanRullen, R., & Thorpe, S. J. (2002) Surfing a spike wave down the ventral stream. *Vision Research*, *42*, 2593-2615.

Walker, S., Stafford, P., & Davis, G. (2008). Ultra-rapid categorization requires visual attention: Scenes with multiple foreground objects. *Journal of Vision*, 8(4):21, 1-12. doi:10.1167/8.4.21.

Walther, B. D., & Fei-Fei, Li. (2007) Task-set switching with natural scenes: Measuring the cost of deploying top-down attention. *Journal of Vision*, *7*, 1-12.

Watamaniuk, S. N. J., Sekuler, R., & Williams, D. W. (1989). Direction perception in complex dynamic displays: The integration of direction information. Vision Research, *29*, 47–59.

Wilson, H. R., & Bergen, J. R. (1979). A four mechanism model for threshold spatial vision. *Vision Research, 19*, 19-32.

Wolfe, J. M. (1999). Inattentional amnesia. In M. Coltheart (Ed.), *Fleeting memories* (pp. 71–94). Cambridge, MA: MIT Press.

Wolfe, J. M., Cave, R. K., & Franzel, L. S. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 419–433. http://dx.doi.org/10.1037/.15.33419.

Appendix A: Example stimuli used in Chapter 2

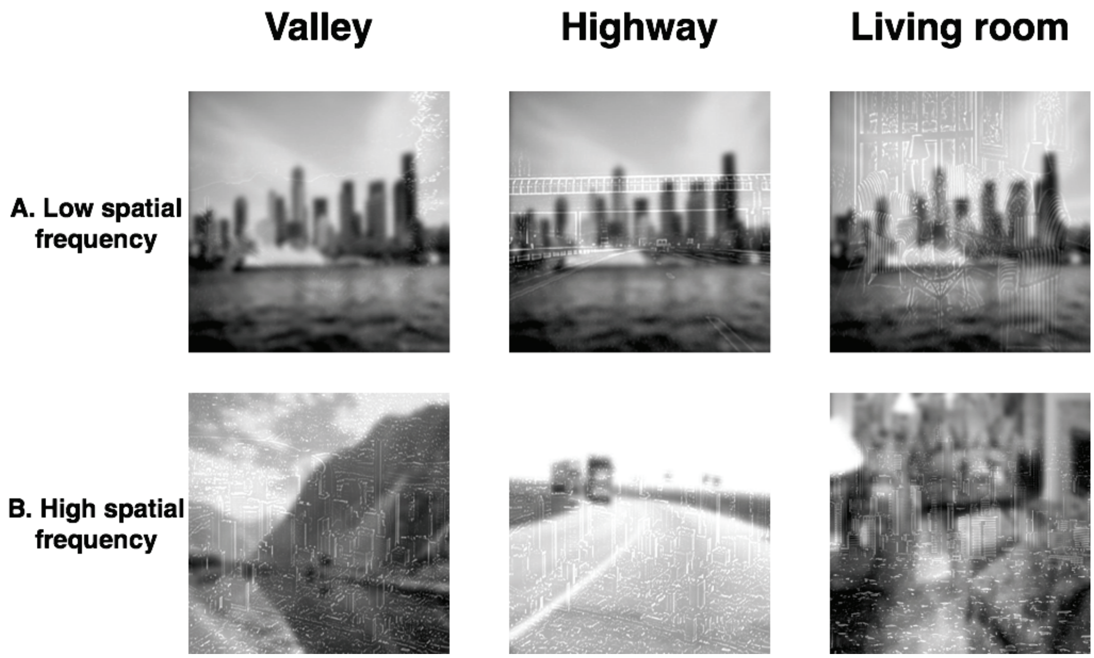


Figure A.1. Examples of hybrid city stimuli. A) A low-pass filtered city scene combined with a high-pass filtered valley, highway, and living room scene, respectively; B) A high-pass filtered city scene combined with a low-pass filtered valley, highway, and living room scene, respectively.

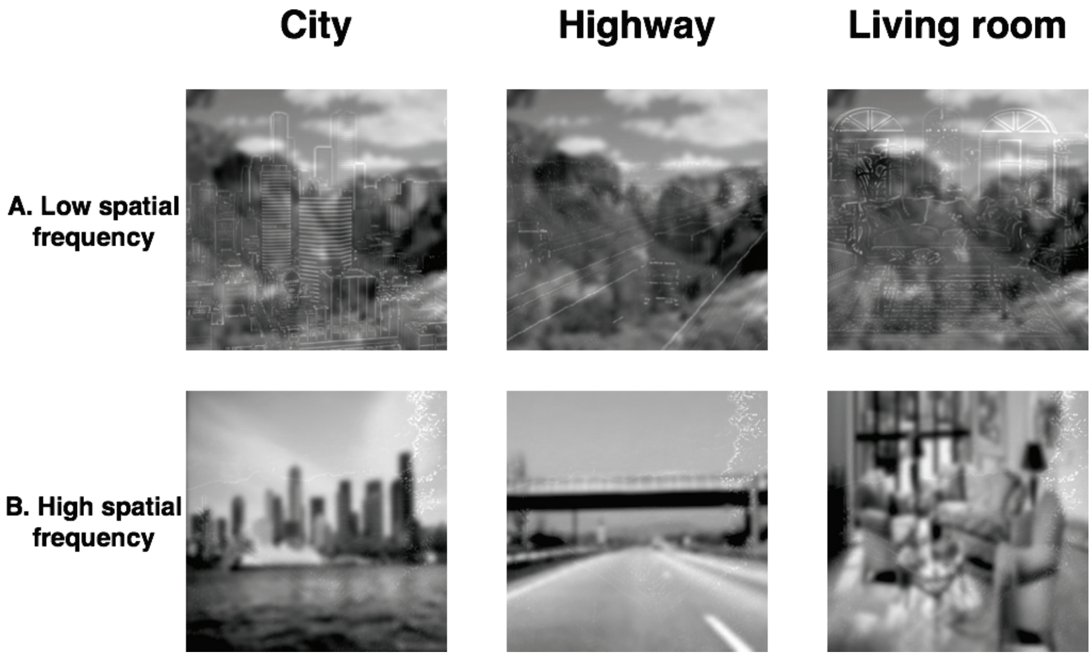


Figure A.2. Examples of hybrid valley stimuli. A) A low-pass filtered valley scene combined with a high-pass filtered city, highway, and living room scene, respectively; B) A high-pass filtered valley scene combined with a low-pass filtered city, highway, and living room scene, respectively.

Figure A.3. Examples of hybrid highway stimuli. A) A low-pass filtered highway scene combined with a high-pass filtered city, valley, and living room scene, respectively; B) A high-pass filtered highway scene combined with a low-pass filtered city, valley, and living room scene, respectively.

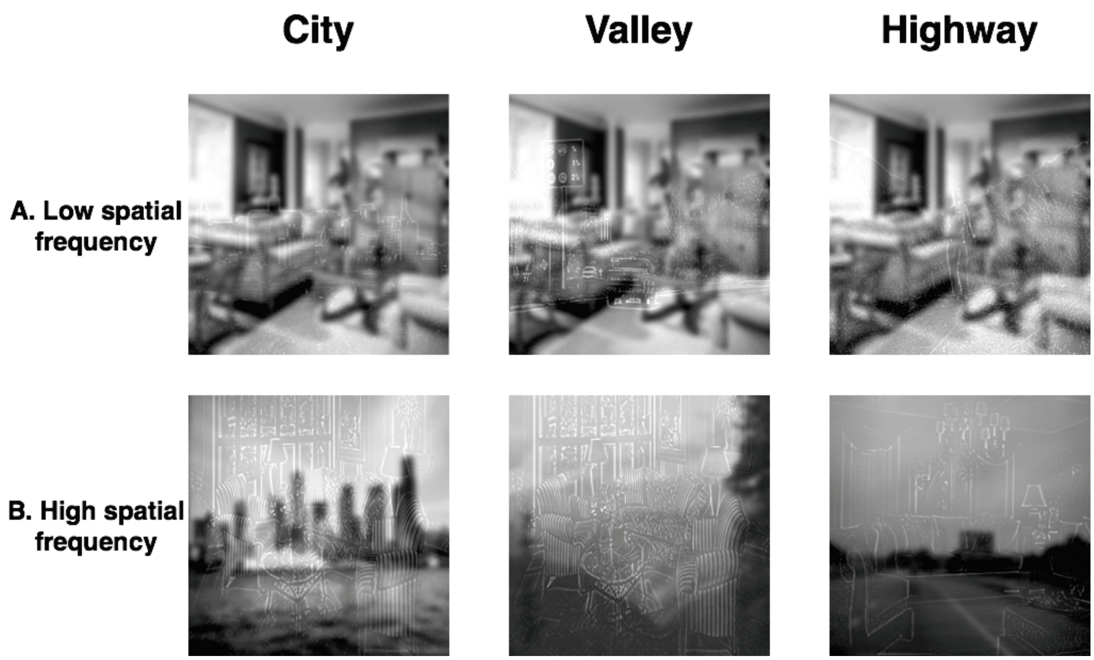


Figure A.4. Examples of hybrid living room stimuli. A) A low-pass filtered living room scene combined with a high-pass filtered city, valley, and highway scene, respectively; B) A high-pass filtered living room scene combined with a low-pass filtered city, valley, and highway scene, respectively.

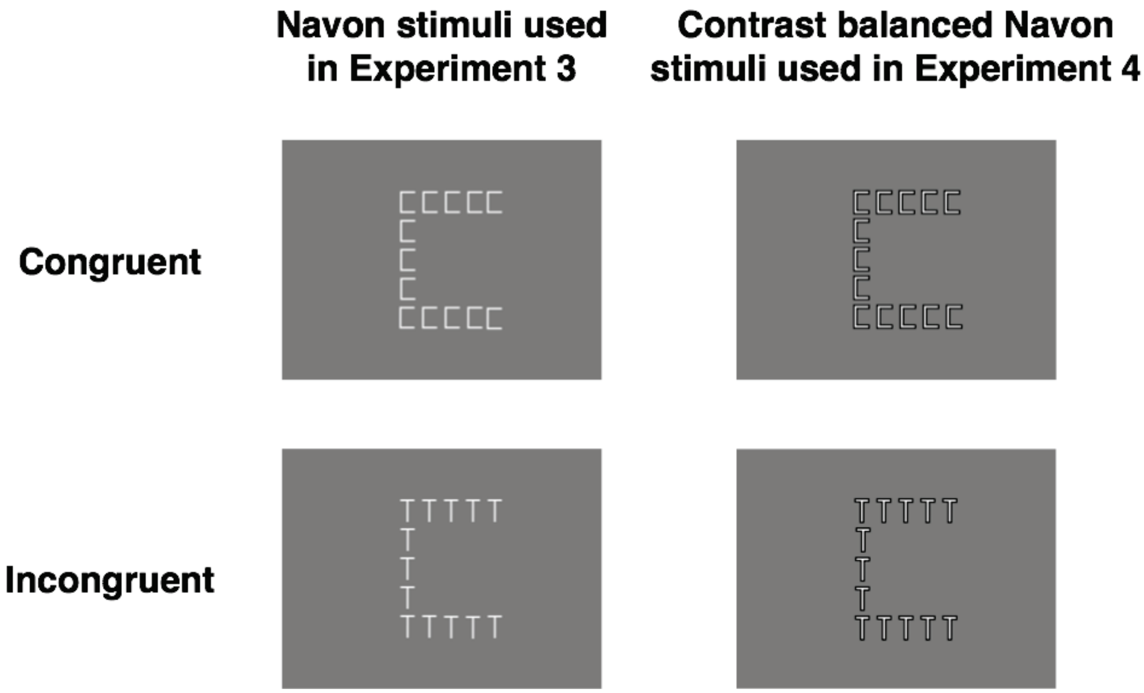


Figure A.5. Examples of congruent and incongruent Navon stimuli used in Experiment 3 and contrast balanced Navon stimuli used in Experiment 4.

Appendix B: Certificate for ethical acceptability for research involving human subjects

CERTIFICATION OF ETHICAL ACCEPTABILITY
FOR RESEARCH INVOLVING HUMAN SUBJECTS

Name of Applicant: Dr Aaron Johnson

Department: Psychology

Agency: NSERC

Title of Project: Visual Processing of Real World Stimuli

Certification Number: 10000119 (UH2006-076-5)

Valid From: September 20, 2012 to: September 19, 2013

The members of the University Human Research Ethics Committee have examined the application for a grant to support the above-named project, and consider the experimental procedures, as outlined by the applicant, to be acceptable on ethical grounds for research involving human subjects.

Dr. James Pfaus, Chair, University Human Research Ethics Committee

01/29/2009