SEMANTIC CONSISTENCY IN BOUNDARY EXTENSION

A Thesis
by
DANIEL CHARLES LACOMBE, JR.

Submitted to the Graduate School
at Appalachian State University
in partial fulfillment of the requirements for the degree of
MASTER OF ARTS

May 2013
Department of Psychology

SEMANTIC CONSISTENCY IN BOUNDARY EXTENSION

A Thesis
by
DANIEL CHARLES LACOMBE JR.
May 2013

APPROVED BY:

_____
Christopher A. Dickinson
Chairperson, Thesis Committee

_____
Kenneth M. Steele
Member, Thesis Committee

_____
Todd McElroy
Member, Thesis Committee

_____
James Denniston
Chairperson, Department of Psychology

_____
Edelma D. Huntley
Dean, Cratis Williams Graduate School

Abstract

SEMANTIC CONSISTENCY IN BOUNDARY EXTENSION

Daniel Charles LaCombe Jr.
B.A., University of Delaware
M.A., Appalachian State University

Chairperson: Christopher A. Dickinson

Two experiments explored the effect of semantic consistency on boundary extension by presenting participants with images depicting scenes with either a semantically consistent or inconsistent object-background relation. These were presented for either a brief or long duration. Based on the existing empirical evidence and a leading theory of scene perception that predicts boundary extension, I hypothesized that there would be more boundary extension for semantically inconsistent scenes compared to semantically consistent scenes for brief durations only. The results of two experiments did not support this hypothesis. I observed no difference in boundary extension for brief image durations in Experiment 1 and a significant difference in boundary extension for long image durations in Experiment 2, with semantically inconsistent scenes eliciting less boundary extension than semantically consistent scenes. These findings are interpreted as evidence for boundary extension being context-independent early in scene perception and context-dependent late in scene perception. I suggest that spatial computation initially operates on object-background spatial relations alone and does not include semantic information, whereas spatial computation later

iv

operates on both object-background spatial relations as well as semantic information about the scene. For semantically inconsistent scenes, the competing semantic information from the context depicted in the background and the context implied by the object interfere with further spatial computation later in scene perception. This explains why there is less boundary extension for semantically inconsistent scenes compared to semantically consistent scenes for long image-durations only. Implications for models of scene perception are also discussed.

Acknowledgments

I would like to acknowledge the support and advice given by my thesis advisor, Dr. Chris Dickinson, as well as that from members of my thesis committee and faculty of the Psychology Department at Appalachian State University. The findings of this thesis were presented at the 2013 North Carolina Cognition Conference with the support of research and travel funding from the Office of Student Research.

Table of Contents

List of Tables

Introduction

Despite the highly variable nature of visual information that we encounter, humans are able to perceive the visual world as a discrete set of representations. In the case of natural scenes, we perceive and identify these visual stimuli effortlessly and in a seemingly automatic manner, regardless of their apparent complexity and ambiguity (Friedman, 1979; Potter, 1976; Schyns & Oliva, 1994). As an example, imagine a blindfolded individual placed in a given kitchen. Previous research suggests that in less than a second after the removal of the blindfold, the individual will know that they are in a kitchen (see Greene & Oliva, 2009; basic-level categorization, Oliva & Torralba, 2001; Potter, 1975, 1976). This rapid identification is accompanied by a wealth of meaningful information that can guide behavior, such as what one might expect to find in this area and where (Biederman, Mezzanotte, & Rabinowitz, 1982; Palmer, 1975; Silva, Groeger, & Bradshaw, 2006). Keeping with the current example, the individual now divested of the blindfold and having identified the surrounding space as a kitchen, would likely have a strong expectation for a stove located on the floor and against a wall, probably beneath a hood fan or cabinets, and adjacent to counter-tops (Biederman et al., 1982; Friedman, 1979). This knowledge could rapidly direct the eyes to this location and allow behaviors for which a stove affords, such as boiling a pot of water (see Henderson & Hollingworth, 1999; Rayner, 2009). Previous research also suggests that this may be accomplished in practically any given kitchen, regardless of idiosyncratic differences, such as slight differences in color, size, and layout of component objects (Friedman, 1979). How does the human visual system accomplish this task of scene perception so rapidly in light of such apparent complexity and variability? How do we use this rapid acquisition of scene identity to facilitate identification of its

constituents? Before these questions can be addressed, we must first consider what scenes are

and the relationships that define them.

Henderson and Hollingworth (1999) have stated that "the concept of a *scene* is

typically defined. . .as a semantically coherent. . .view of a real-world environment

comprising background elements and multiple discrete objects arranged in a spatially

licensed manner" (p. 244). Returning to the kitchen example, a scene of this particular type

would likely include background elements such as vinyl flooring, counter-tops and cabinets,

and may contain objects that are semantically consistent with a kitchen, such as a

refrigerator, a stove, and a sink. These kitchen elements also occur in a spatially licensed

manner, with vinyl flooring being on the floor, not on the ceiling. These constituents of a

kitchen scene and their relationships are presumably derived from prior experiences with

kitchens, which yield an internal model of what a kitchen means, as well as the elements that

are associated with that construct and their relations. The definition given by Henderson and

Hollingworth (1999) implies that scene identification involves the identification of the

separate elements of the scene and their relations, both spatial and semantic. Given that scene

identification could involve the identification of both objects and background elements,

might they influence the perception of one another?

This question has been addressed extensively within the scene perception literature.

The central idea surrounding this empirical enterprise has been that of *scene semantics*, that

is, the semantic relationship between a scene's background and its constituent objects. Early

research in scene semantics has described and examined the contributions of multiple object-

background relations and relational violations in scene perception, including support,

interposition, probability, position, and size (e.g., Biederman et al., 1982). The findings of

this early research suggested that each of these violations had a detrimental impact on scene

perception (i.e., impaired people's ability to identify objects in scenes), with combinations

having a more severe influence than those occurring in isolation. The subsequent literature in

this area of research has primarily addressed violations of probability and location, to which I

refer as semantic and spatial, respectively. Of these two, this thesis will be almost exclusively

concerned with semantic object-background relations, that is, the probability that an object

will co-occur with a background. Once again returning to the kitchen example, a stove can be

considered to be semantically consistent with a kitchen background, whereas a bed would be

considered as semantically inconsistent with a kitchen background. It is important to note

that semantic consistency/inconsistency is used interchangeably in the reviewed literature

with uninformative/informative, congruent/incongruent, and episodic consistency/episodic

inconsistency.

As previously mentioned, scene perception is astonishingly rapid, requiring an image

presentation of only about 100 ms for people to identify a scene's basic-level category (e.g.,

kitchen; Potter, 1975, 1976; Schyns & Oliva, 1994). How can such rapid scene identification

be explained? The most prominent theoretic proposition to explain this phenomenon, of

which has received due empirical attention, is the concept of a perceptual schema

(Biederman, 1981; Friedman, 1979). The concept of a perceptual schema is an old one,

dating back even as far as early philosophy. A bit closer to home, the idea of a perceptual

schema, or an internal model, has been widely recognized as a necessity in the field of

artificial intelligence to disambiguate deprived visual input (e.g., incomplete objects; see

Friedman, 1979). The same seems to apply for humans (as frame theorists have posited;

Biederman, 1981; Friedman, 1979); over the course of cognitive development, humans

develop schemas that facilitate perception in an automatic manner (Friedman, 1979). This theory, which will be referred to as the perceptual schema model, posits that global relations between backgrounds and objects are extracted rapidly due to expectations derived from previous experiences and that their probabilistic co-occurrence facilitates perception when they are presented simultaneously. Objects and backgrounds that have not co-occurred often, or at all, in the subject's experience require more perceptual effort to extract local visual detail in order to accomplish identification. In short, the visual system has developed an internal model of the external world, upon which visual information is mapped.

The earliest sources of evidence for this type of model resulted from a series of experiments (Biederman 1972; Biederman, Glass, & Stacy, 1973; Biederman, Rabinowitz, Glass, & Stacy, 1974) in which global relations of a scene were manipulated. This was accomplished by jumbling pictures, which specifically means segregating an image into a given number of squares (in this case six) and reassembling them without rotating them. The participant's task was to identify a target object at a cued location from four alternative objects that all appeared within the scene. The square containing the target object was never moved so that the normal spatial position of the object was held constant while the surrounding context was manipulated. If global context surrounding the object facilitates object identification, then one would predict better performance in the non-jumbled condition. Indeed, these studies found an object identification advantage for targets in the non-jumbled condition over targets contained in jumbled pictures. This was taken as evidence for the rapid acquisition of a perceptual schema via global relations that facilitated object recognition; those global relations were deteriorated in jumbled pictures, hence poorer object recognition. These seminal studies should be taken with a grain of salt, however, as

criticisms of this early paradigm may constrain the validity of their findings (such as

introducing new contours by dividing the image – possibly the source of perceptual

difficulty; see Henderson & Hollingworth, 1999), some of which are discussed later.

Since these initial sources of evidence, there have been many empirical studies aimed

at testing the predictions of the perceptual schema model as it relates to scene semantics. In

behavioral studies, measures of object detection in experiments in which the semantic

consistency between objects and backgrounds are manipulated have been used as evidence of

a perceptual schema. In these paradigms, a scene is presented briefly and then masked (i.e.,

interrupted by a homogeneous array so that visual information cannot be held in sensory

memory). Next, the participant is prompted to determine whether a given target appeared at a

cued location within the scene. Measurement of detection in these studies is operationalized

as a sensitivity measure, *d'*. The logic behind the object-detection approach is that if a

perceptual schema is activated rapidly, then objects consistent with a particular schema will

be more readily identified. This is the primary paradigm used in the scene semantics

literature; earlier studies have used priming and change detection paradigms and will be

described shortly. In eye movement paradigms, which are not mutually exclusive with

object-detection approaches, eye trackers record where people are looking while exploring a

scene. Information is extracted almost exclusively during fixations because visual

information is suppressed during saccades (see Irwin & Brockmole, 2004; Matin, 1974).

Thus, the duration of a fixation is taken to be a measure of the amount of processing required

for an object at that given fixation; longer fixations reflect more perceptual processing and

extraction of local visual information. These two behavioral measures represent the majority

of support for perceptual schema models of scene perception; therefore, they will be the two

primary measures reviewed here. In addition to these sources of evidence, modifications of

the object identification paradigm, event-related potential measurements, and computational

modeling will be reviewed briefly.

Following the first sources of evidence for contextual facilitation from jumbled-scene

paradigms (Biederman 1972; Biederman et al., 1973, 1974), Palmer (1975) asked whether

scenes would contextually prime certain objects that did not appear in the scene. To do this,

he developed a priming paradigm in which a line drawing of a scene (e.g., a counter-top with

a cutting board) was presented for 2 s followed by a target object (e.g., loaf of bread)

presented for durations of 20, 40, 60, or 120 ms. After the object presentation, the subject

was then instructed to write down the name of the object. The results of the study showed a

clear advantage for identifying objects that were consistent with the context conveyed by the

preceding image, for all durations. These results suggest that the prime image activated a

perceptual schema which contained candidate objects that have occurred in that context in

previous experiences.

Friedman (1979) asked whether a perceptual schema might make the perception of

objects in prototypical scenes an automatic and top-down process compared to scenes that

had unlikely objects presented within them, which may be characterized by more controlled

extraction of perceptual information. The experimental paradigm – essentially a change

detection paradigm – involved a presentation of two pictures separated by a mask in which a

single object changed. The change was a token change (e.g., a lamp exchanged for another

lamp, with only subordinate differences), a location change, a deletion, or a type change (of

differing semantic consistency; e.g., high to low – farm scene exchanging cow for

hippopotamus). Subjects were asked to indicate if a change was detected between the two

pictures. She found that subjects noticed more changes from semantically consistent to

inconsistent objects, whereas substituting a consistent object with another consistent object

was not detected as accurately. The eye movement data from this study were complementary

to the object change detection task in that consistent objects were fixated for shorter

durations than inconsistent objects (similar results to Loftus & Mackworth, 1978).

Collectively, these results suggest that objects consistent with a scene were processed more

automatically, requiring less encoding of perceptual information for identification and

therefore resulting in poorer discrimination in the change detection task (Friedman, 1979);

therefore, an implication for the perceptual schema theory may be that objects inconsistent

with a background will recruit more attention to extract local visual detail. This point will be

discussed more in the next section.

The study by Friedman (1979) addressed exclusively the question of probabilistic

relations between scene context and the objects occurring within them. Biederman et al.

(1982) extended this work by asking whether both probabilistic relations and other semantic

and syntactical relations, such as size, support, interposition, and position affected the

perception of objects within a scene. In this experiment, the name of the target object was

presented until the participant self-initiated a given trial. Once initiated, a scene would appear

for 150 ms followed by a spatial cue at which time the object to be detected would appear.

Following this, two object alternatives were provided – the object that was present, and

another object that was not present. The target object in a given trial could either be

consistent with the context, or inconsistent, violating semantic (e.g., probability of co-

occurrence) or syntactic (e.g., physical support) relations, or compound violations (e.g., an

object that is both improbable and not physically supported). Object detection was

operationalized as a sensitivity measure, *d'*. Biederman et al. (1982) hypothesized that if a

scene schema is rapidly activated, then objects that are consistent with the expectations

generated by that schema should be detected more accurately, as indicated by the sensitivity

measure, *d'*. The results of the study supported this hypothesis, with objects violating

semantic relations having lower sensitivity measures, and with compound violations having

even a larger decrement.

Despite these sources of support for the perceptual schema model of scene

perception, there also exists evidence contrary to this model. De Graef, Christiaens, and

d'Ydewalle (1990) noted an alternative interpretation of the results generated by the object

detection paradigm used by Biederman et al. (1982). Whereas the object detection advantage

was interpreted by Biederman et al. (1982) as increased perceptibility of objects that were

consistent with a perceptual schema, de Graef et al. (1990) suggested that this advantage

could also be explained by post-perceptual guessing that placed an overreliance on contextual

information. That is, instead of more accurately perceiving an object that is consistent with a

particular perceptual schema, the contextual information extracted from the 150 ms stimulus

duration (e.g., Schyns & Oliva, 1994) would be available after stimulus offset, allowing

participants to make educated guesses as to whether the object was present or not, even if the

object had not been perceived. To address these alternatives, de Graef et al. (1990) developed

a paradigm in which the task minimized the explicit use of contextual information by

instructing participants to engage in an object-decision task (indicating if closed entities were

known objects) and by using first-fixations of objects as a more unobtrusive measure of

object identification (unreflective of post-perceptual processes). Line drawings of real-world

setting were presented that either adhered to or violated some of the five semantic violations

described by Biederman et al. (1982) and that contained different amounts of non-object

targets (zero to three). Once fixation was made to a central fixation cross, a given scene was

presented for 8 s, after which participants were instructed to press a key once for every non-

object present in the scene (data of which was not actually recorded). Main effects of longer

first-fixations on display items were found for semantic violations, consistent with earlier

studies (Friedman, 1979; Loftus & Mackworth, 1978); however, the authors performed

further analyses by using the median fixation count (eight fixations) as a cutoff for first-

fixation durations to create early (first eight fixations) and late (n fixations after eighth

fixation) fixation-moment groups. This analysis revealed a main effect of longer-fixation

durations in the late fixation-moment group, indicating that differences in first-fixation

durations between semantically consistent and inconsistent objects only emerged late in

scene viewing, which is inconsistent with a rapidly activated perceptual schema (Biederman

et al., 1982).

Hollingworth and Henderson (1998; 1999) also addressed some of the

methodological issues present in the Biederman et al. (1982) research paradigm, which has

been considered as the most convincing source of evidence for the perceptual schema model

(as noted by Henderson & Hollingworth, 1999). They noted that the sensitivity measure used

by Biederman et al. (1982) calculated false alarm rates by averaging across both consistent

and inconsistent object conditions, which may have artificially raised the sensitivity measure

for the base (consistent) condition compared to violation (inconsistent) conditions. That is,

false-alarm rates were higher in the base conditions; by averaging with false-alarm rates from

the violation condition, this lowers the overall false-alarm rate value. When d' is calculated,

this lower overall false-alarm value artificially raises base rate values and lowers violation

values, which may have been the source of significant differences. They also noted that the

target label prior to scene viewing may have been used to guide search and that the spatial

cue may have served as evidence for the types of objects that likely occur at that location

(Henderson, 1992; Hollingworth & Henderson, 1998). Lastly, they noted (as did Boyce &

Pollastek, 1992; de Graef et al., 1990) that there may have been a response bias associated

with an overreliance on contextual information.

        To address the bias in the detection sensitivity measure, Hollingworth and Henderson

(1998) calculated detection sensitivity by calculating false-alarm rates for each condition

separately; for both conditions, sensitivity was based only on the correct detection of a

particular object when it was present and false detection of the same object when it was not

present (Hollingworth & Henderson, 1998). The target preview/location cue advantages were

addressed by placing the object label after the scene presentation, without a location cue.

This involved modifying the Biederman et al. (1982) paradigm by having participants fixate

a central fixation cross for 500 ms, presenting a series of Xs for 1.5 s (which was used to

equate stimulus presentations of the original paradigm) followed by scene presentation for

200 ms. and then presenting the object label embedded with a pattern mask until response

(i.e., participant indicating the presence or absence of the object label by responding yes or

no). Finally, the response bias was controlled for by using a forced go/no-go paradigm in

which both object labels are either semantically consistent or inconsistent. Also involving a

modification of the Biederman et al. (1982) paradigm, this was implemented in a separate

experiment by having participants fixate at a central fixation cross for 500 ms, presenting the

scene for 250 ms followed by a 30 ms mask, and lastly having two object labels – that were

either both semantically consistent or inconsistent – presented beside each other until

response (i.e., pressing either a left or right button to indicate if they thought that object label

on the left or right appeared in the scene, respectively). When all of these biases were

separately controlled, the consistent object performance advantage disappeared, or even

reversed in some cases (label post-view). This difference in findings has motivated the

development of a functional isolation model (FIM) of object and background perception

(Henderson, 1992; also see Henderson & Hollingworth, 1999), which the space allotted here

does not allow for ample review, although in brief, it posits that the perception of objects and

the backgrounds in which they are presented do not influence one another early in scene

perception.

Finally, Henderson, Weeks, and Hollingworth (1999) asked if the discrepant findings

of earlier eye-tracking studies in the semantic consistency literature could be attributable to

differences in paradigms used. In this article, they noted that a potential source of difference

between the de Graef et al. (1990) and Loftus and Mackworth (1978) studies could be the

nature of the task, with the former being a visual search task and the latter being a memory

test. To ensure that the absence of an early contextual effect was not attributable to the type

of the task, Henderson et al. (1999) used line-drawing stimuli similar to the two previous

studies, but used a memory task akin to Loftus and Mackworth (1978), although in both

cases, the memory test never took place. The paradigm involved a 15-s presentation of line-

drawing scenes in which either semantically consistent or inconsistent objects appeared, and

the participant was instructed to prepare for a memory test after all scenes were presented

(which, as mentioned previously, never took place). Supporting the findings of de Graef et

al. (1990), the results indicated that initial fixation placement was not controlled by the

semantics of a scene, once again contradicting the predictions of the perceptual schema

model that posits rapid activation upon viewing. Instead it appeared that target objects were

fixated when approximately 3° away from the current fixation location, suggesting that

fixations are controlled more parafoveally (i.e., area of the retina between fovea and

periphery) than by context, as suggested by Henderson, Pollatsek, and Rayner (1989), and

also supported the proposed FIM (Henderson, 1992). Although this was the case, inconsistent

objects were nonetheless fixated for longer durations on the first-pass fixation.

The research reviewed so far has displayed a transition from initial strength of the

perceptual schema model of scene perception to ever growing skepticism of its validity;

however, more recent studies have provided support for the concept of an internal model.

Specifically, Davenport and Potter (2004) asked the question of whether objects and

backgrounds will influence the identification of one another based on their semantic,

probabilistic relations, thus further testing the predictions of the perceptual schema model.

The approach they used to address this question paralleled those used in previous studies, but

differed in subtle ways. The stimuli used in their study were real, color photographs as

opposed to line-drawings and contained a single foreground object instead of multiple

objects. Additionally, they used an identification task (i.e., participants were asked to name

the object, the background, or both) whereas Henderson and Hollingworth (1998, 1999) used

a forced-choice two-alternative task which forced participants to guess if they were unsure,

and may have given inconsistent objects an asymmetrical advantage (see Davenport &

Potter, 2004). As in previous studies, the hypothesis tested, which is generated by the

perceptual schema theory, was that context should be activated rapidly enough to influence

the perception of elements within a scene. The results supported this hypothesis, finding an

object identification advantage when objects were presented within a semantically consistent

background for 80 ms. These findings contradict earlier studies providing evidence against

the existence of a rapidly activated perceptual schema but the authors noted that these

differences are likely attributable to differences in paradigms; the stimuli and task used by

Davenport and Potter (2004) were probably more sensitive to contextual effects elicited by

the proposed perceptual schema, and those effects may have been overshadowed in more

complex stimuli used by Henderson and Hollingworth (1998, 1999).

   In support of this newfound evidence for the existence of a rapidly activated

perceptual schema, Bonitz and Gordon (2008) provided eye movement results that support

the perceptual schema model of scene perception. As with the previous eye tracking studies

reviewed here, this study was concerned with fixation durations as a measure of perceptual

effort required to encode an object at fixation. The paradigm used real color photographs

presented for 10 s during which participants were instructed to explore the scene freely and

were subsequently asked to rate the scene for pleasantness (this was done to engage

participants in viewing the scene when eye movements were being recorded). Results of the

study revealed that fixation durations were longer for inconsistent objects. These findings are

consistent with earlier eye tracking studies (Friedman, 1979; Loftus & Mackworth, 1978) but

are inconsistent with more recent studies (de Graef et al., 1990; Henderson & Hollingworth,

1998, 1999). The authors attributed the discrepancy in results to the nature of the stimuli; like

the explanation given by Davenport and Potter (2004), Bonitz and Gordon (2008) suggested

that their more simplistic stimuli may have been more sensitive to perceptual effects of

semantic inconsistency, whereas the more complex stimuli used by de Graef et al. (1990) and

Henderson and Hollingworth (1998, 1999) may have made it more difficult for participants

to identify objects thus not allowing perceptual effects to emerge in those experimental

paradigms.

More recently, Sun, Simon-Dack, Gordon, and Teder (2011) have also provided

supporting evidence of the theory of a rapidly activated perceptual schema that facilitates

object perception. Like many of the previous studies, they asked whether scene context

would facilitate perception of the constituent objects. The paradigm they used to answer this

question was a go/no-go task in which participants were instructed to indicate whether an

animal appeared within a non-masked scene presented for 20 ms. To manipulate the

influence of scene context, they either retained, deleted, or phase randomized the background

in the stimuli surrounding a foreground object. Phase randomization involves a procedure

where random noise can be introduced that effectively removes semantic information while

preserving low-level image properties, such as luminance and spatial frequency; this allowed

any differences under these conditions to be attributable to loss of semantic information and

not differences in low-level image properties. Behavioral results indicated that accuracy was

higher and reaction times were faster in the animal present go/no-go task with the

background intact compared to both the deleted and phase randomized background

conditions. The authors of this study also collected event-related potentials (ERPs) while

participants completed the task. ERPs are graded levels of positive or negative electrical

activity measured at the scalp that are indicative of brain processes that occur at predictable

stages in time after the onset of an event (i.e., latency of an ERP component, which can be

used to provide estimates of visual processing over time; Thorpe, Fize, & Marlot, 1996).

Consistent with the reaction time results, the onset latency of the frontal component

associated with the go/no-go task (e.g., Thorpe et al., 1996) was delayed for objects

occurring outside of their original context, providing converging evidence that scene context facilitates object perception.

Up to this point, the way in which the literature has been reviewed has focused almost exclusively on a perceptual schema facilitating subsequent analysis of objects within a scene. But could the contextual elicitations of objects influence the perception of backgrounds? In addition to finding an identification advantage for objects occurring within semantically consistent backgrounds, Davenport and Potter (2004) and Davenport (2007) provided the first sources of evidence for identification advantages for backgrounds in which semantically consistent objects were presented. When asked to identify the background of a masked image presented for 80 ms, accuracy was significantly higher when a foreground object was consistent with the background than when it was inconsistent. This finding led to the proposal of an interactive model of scene perception, which posits that scenes and their constituent elements are processed holistically; objects and backgrounds provide contexts for one another in a mutually constraining manner. Although not a major departure from the perceptual schema model of scene perception, the interactive model does not assume that a schema is necessarily activated prior to object context, but rather that they occur in parallel; "Objects provide the context for the background, and the background provides the context for foreground objects" (Davenport, 2007, p. 394). Following this theoretical proposal, a couple of noteworthy studies have provided supporting evidence for this theory, arriving in the form of behavioral and computational model findings.

Although they did not explicitly manipulate object-background semantic consistency, Joubert, Rousselet, Fize, and Fabre-Thorpe (2007) provided supporting evidence for the influence of object semantic consistency on background perception. The authors asked how

quickly humans can globally contextualize scenes at the superordinate level, such as whether they were man-made or natural, which is more general than basic-level contextual identification (e.g., mountain, valley; Oliva & Torralba, 2001). To do so, they asked participants to complete a go/no-go task in which they categorized a real color photograph of either a man-made environment or a natural environment presented for 26 ms. These scenes were accurately categorized at 96%, with reaction times around 390 ms. When analyzed post-hoc for object congruency (e.g., man-made object such as a building in a city would be congruent, while a tree in a city would be incongruent), results indicated that scene categorization was significantly impaired for incongruent scenes. An effect of object inconsistency interfering with scene categorization at such an early time interval provides supporting evidence for the interactive model of scene perception; rapid object recognition may be available early enough to influence scene identification and may not necessarily occur after schema activation.

The findings of Joubert et al. (2007) discussed above were replicated by Mack and Palmeri (2010), who also used a go/no-go categorization task of either man-made environments or natural environments using color photographs but instead had an explicit manipulation of object-background consistency. As in Joubert et al. (2007), categorization performance was diminished for inconsistent trials relative to consistent trials, and reaction times were faster for consistent object-background conditions. Mack and Palmeri (2010) then furthered these findings by developing a computational model that integrated a scene categorization model (Oliva & Torralba, 2001) and a perceptual decision making model (Ratcliff, 1978) to simulate behavioral data. The scene categorization model developed by Oliva and Torralba (2001) extracts the global spatial structure of a scene and compares this to

derived spectral properties that were diagnostic of particular spatial characteristics (e.g.,

naturalness) – a feature space termed the *Spatial Envelope*. In the case of an inconsistent

scene (e.g., forest scene with a shed as a foreground object), the spatial statistics will be

slightly discrepant from those in a consistent scene (e.g., a forest scene with a bush as a

foreground object) and therefore provide lower quality evidence towards its correct

categorization (i.e., spatial statistics further from those diagnostic of either natural or man-

made scenes). When conjoined with the decision making model (Ratcliff, 1978), poorer

perceptual evidence results in slower reaction times and lower accuracy in scene

categorization. Consistent with the behavioral data, the model's simulation revealed that

accuracy in categorization was higher and reaction times were faster for semantically

consistent scenes. The findings of this study imply that slight differences in the overall

spatial statistics of scenes (e.g., right-angles and straight horizontal/vertical lines in man-

made scenes and oblique contours and curved lines in natural scenes) could explain the

perceptual differences between consistent and inconsistent scenes without stipulating distinct

object and background recognition processes.

It should be clear by now that there exists evidence that semantic consistency,

specifically the probability that an object will co-occur with a background, has an impact on

scene identification. Although there exists mixed evidence within the literature for the effects

of scene semantics, these discrepancies can be explained when the differences in stimuli are

considered; experimental paradigms that find evidence for semantic inconsistency

detrimentally impacting scene perception use more simplistic stimuli that may be more

sensitive to these types of perceptual effects. Having said this, the plausibility of perceptual

schema and interactive models of scene perception remain strong despite some sources of

contradicting evidence, yet further research remains to be implemented for these theories to become more firmly supported. Nonetheless, it is safe to say that there is evidence that semantic consistency does appreciably affect the perception of scenes and their constituent elements. Scene identification, however, is only one aspect of scene perception. Could semantic consistency affect other aspects of scene perception?

Semantic Consistency in Boundary Extension

Identifying and categorizing scenes is a fundamental aspect of scene perception; it allows one to understand what objects and events are likely to take place within that space as well as the types of behaviors that type of space affords. Another aspect of scene perception involves the representation of the spatial expanse depicted in a view of a scene (e.g., a close-up or a wide-angle view). The importance of such a computation is illuminated when considering the constraints of the visual system. Because of the physical structure of the eye and their socket positions in the front of the head (for humans at least) the entire visual world can never be completely accessed at any given point in time. Additionally, information received by the retina is not homogeneously represented throughout the visual system; input at the fovea is of highest acuity, with resolution falling as a function of distance from the fovea (see O'Regan, 1992); therefore, visual input in the periphery is of low resolution and often considered as "fuzzy." Because we can never simultaneously view the surrounding space at once and that even a given view of the world is of low quality at the periphery, the need for the visual system to be able to understand the space depicted within a scene becomes clearer; that is, we often need to reference or orient to objects and events occurring in the spatial surround (which are not completely available in the present view). It is therefore advantageous to predict the nature of that area (e.g., layout, surfaces, objects, etc.).

Boundary extension (Intraub & Richardson, 1989), which is the phenomenon of remembering a more spatially expansive view than was originally shown, presumably reflects such a computation. The typical boundary extension paradigm involves the presentation of either a close-up or wide-angle view of a scene followed by a retention interval and then a memory test, typically a recognition task, in which the test view can either be identical to the stimulus or different (i.e., more close-up or more wide-angle). When presented with identical stimulus and test views, participants on average tend to report the test picture as being closer-up in comparison to the stimulus picture, with the effect being more pronounced for close-up views than for wide-angle views, and with the latter sometimes not eliciting unidirectional memory distortions (average response of "same view") depending on how wide-angle the view is (e.g., Dickinson & Intraub, 2008; Intraub, Bender, & Mangels, 1992; Intraub, Daniels, Horowitz, & Wolfe, 2008; Intraub & Dickinson, 2008; Intraub & Richardson, 1989; Park, Intraub, Yi, Widders, & Chun, 2007). The predominant interpretation of participants tending to report test views as closer-up is that stimulus views are remembered as being more wide-angle than they appeared.

This interpretation is corroborated by the asymmetry in average responses made when the test view is an alternative; when the stimulus is a close-up version of a picture and the test image is the wide-angle version, the test image is rated as more similar to the stimulus than when the stimulus is the wide-angle version of a picture and the test image is the close-up version (e.g., Intraub et al., 1992). In both cases, the same close-up and wide-angle versions are being compared, with the only difference being which one serves as the stimulus image and which serves as the test image. This asymmetry provides converging evidence for the interpretation stated above; close-up stimuli elicit boundary extension, making the memory

representation of a given image closer to the wide-angle version, whereas wide-angle stimuli

elicit little or no boundary extension, thus preserving the difference between the views. Also

parallel to these findings and interpretations, Park et al. (2007) found selective attenuation

(neuronal behavior that indexes the sequential presentation of identical stimuli) in neural

nuclei known to be involved in scene perception (i.e., the parahippocampal place area and the

retrosplenial cortex) only when a close-up stimulus was followed by a wide-angle test view,

and not vice-versa. Like the behavioral results above (Intraub et al., 1992), this brain area

seems to respond to wider-angle test pictures as if they were more similar to the closer-up

stimulus picture presented earlier.

There is evidence that this error of commission appears early in scene perception

(Dickinson & Intraub, 2008; Intraub & Dickinson, 2008). Intraub and Dickinson (2008) used

the typical boundary extension paradigm described above, in which an image is presented,

masked, and followed by a recognition test where either an identical or alternative view is

shown. In this experiment, images were presented for 250 ms and masked at the shortest time

interval of 42 ms, commensurate to the duration of a saccade (Rayner, 2009). Consistent with

previous studies, a visual mask intervening between stimulus and test views – in this case as

brief as 42 ms – was sufficient to elicit memory errors diagnostic of boundary extension (i.e.,

the behavioral asymmetry described above as well as the tendency to rate close-up test

pictures as more close up than the close-up stimulus of the same view as the test picture).

Intraub and Dickinson (2008) suggested that it was highly unlikely that boundary extension

could be occurring during the interval of a 42 ms mask, and instead suggested that it may be

a result of processes occurring during scene viewing, rather than after.

The claim that the processes eventually resulting in boundary extension occur, or at least began, during scene viewing is central to Intraub's (2010) *multi-source model* of scene perception. The multi-source model (Intraub, 2010) posits that scene perception is essentially an act of spatial cognition; when we are looking at a part of the world, visual information is mapped onto an egocentric spatial-framework that surrounds the viewer and amodal input (i.e., non-perceptual; Kanizsa, 1979; Yin, Kellman, & Shipley, 2000) fills in low-resolution and incomplete information at the edges of the view. When viewing a scene, these multiple sources are simultaneously involved and are reliably distinguished given that the view is available. When a scene is removed or interrupted, however, sources of information are no longer reliably demarcated, as they are hypothesized not to have explicit "tags" associated with them. Under these circumstances, boundary extension effectively becomes a *source-monitoring error* (Johnson, Hashtroudi, & Lindsay, 1993); amodal input is erroneously attributed as visually derived, which predicts a bias to remember views as more wide-angle – consistent with boundary extension. This model provides a more parsimonious explanation of the findings, especially when compared to attempts at modifying existing models of visual memory buffers (e.g., transsaccadic memory, Irwin, 1991; visual short-term memory, Phillips, 1974; conceptual short-term memory, Potter, 1976) by assuming additional perceptual mechanisms such as an "extrapolation process" (Intraub & Dickinson, 2008).

As mentioned previously, boundary extension appears to reflect the computation of mapping a discrete visual input onto a continuous spatial framework, as proposed by the multi-source model (Intraub, 2010). Indeed, studies have shown that boundary extension occurs only for scenes that depict partial views of a larger surrounding space (Gottesman & Intraub, 2002; Intraub, Gottesman, & Bills, 1998). Intraub et al. (1998) presented individuals

with a sequence of images followed by a recognition test in which the test view was either

identical (e.g., close-up stimulus, close-up test view; wide-angle stimulus, wide-angle test-

view) or an alternative (e.g., close-up stimulus, wide-angle test view; wide-angle stimulus,

close-up test view), with a given image having both close-up and wide-angle versions. They

manipulated what was depicted in a scene by creating line drawings of objects from intact

images and placing them on blank backgrounds; line drawings of objects on blank

backgrounds depicted a "complete" view (i.e., no surrounding space beyond edges of view),

whereas intact images with backgrounds depicted a partial view (i.e., surrounding space

beyond edges of view). The intact pictures elicited patterns of memory errors diagnostic of

boundary extension (i.e., boundary extension for close-up images and no bias towards

unidirectional errors for wide-angle images), whereas pictures containing an object on a

blank background did not (i.e., extension for close-up images and restriction for wide-angle

images – a pattern of results characteristic of normalization, in which errors occur in

direction of the average view size). Thus, it appeared that the visual system was not treating

these two types of stimuli equivalently. For those stimuli containing just an object on a blank

background, however, patterns of memory errors diagnostic of boundary extension occurred

when participants were asked to imagine the object as being part of a more spatially

expansive view, which the authors suggested supported the existence of a perceptual schema

that could be activated via a scene depicting a larger space or via imagination. This basic

pattern of results has been replicated by Gottesman and Intraub (2002) in which background

construal was manipulated by showing a participant the object being placed on a blank

background. Under these conditions, boundary extension was eliminated, supporting the idea

that boundary extension is sensitive to whether we construe a scene as depicting a larger

space.

Intraub et al. (1998) suggested that the reason objects on blank backgrounds did not

elicit patterns of memory errors indicative of boundary extension was because these types of

stimuli violated all five semantic relationships proposed by Biederman et al. (1982; i.e.,

probability, position, interposition, support, and size). That is, images violating all semantic

and syntactical relationships that define a real-world scene will not depict a larger spatial

surrounding and therefore should not induce memory errors diagnostic of boundary

extension. They also noted that in boundary extension experiments using semantically

inconsistent objects and backgrounds (e.g., stuffed animal on stairs, banana on rocks),

boundary extension was not eliminated. How might object-background semantic consistency

influence boundary extension? Previous research in the scene semantics literature provides

supporting evidence for the claim that when objects are inconsistent with the surrounding

background, more attention is required to process them fully; if this were the case, it would

predict more boundary extension for inconsistent scenes than consistent scenes. The

following discussion briefly reviews literature providing supporting evidence for this

attention-based prediction.

One source of evidence for increased attentional allocation to inconsistent objects

compared to consistent objects is provided by the results from the change-detection paradigm

used by Friedman (1979), in which substitutions of consistent objects with another consistent

object were detected with lower accuracy than substitutions of consistent objects with

inconsistent objects (or vice versa). This finding suggests that more attention was devoted to

the perception of inconsistent objects, allowing participants to extract more local visual detail

that served to aid discrimination between changes across pictures when inconsistent objects were involved. Another source of support for the hypothesis that inconsistent objects are likely to receive more attention than consistent objects is provided by the results of eye tracking studies that report longer durations for initial fixations to inconsistent objects compared to consistent objects (Bonitz & Gordon, 2008; Friedman, 1979; Loftus & Mackworth, 1978). Together, these sources of evidence suggest that inconsistent objects recruit more attention compared to consistent objects. If this is the case, it may be possible that allocating more attentional resources to a central foreground object in a scene would make it more difficult to delineate between perceptual and amodal sources of information at the borders of a view during a memory test, resulting in larger amounts of boundary extension.

Indeed, it has recently been proposed that attention may serve to delineate between sources of information involved in scene perception (multi-source model; Intraub, 2010), based on the findings of Intraub et al. (2008). In this study, images were presented for 750 ms with 2s and 5s superimposed upon the images and a view-recognition test was given on each trial for both the memory-only and divided-attention conditions. In the divided-attention condition, participants were instructed to prioritize the counting of the number of 5s superimposed on the image and to indicate their response immediately after image offset when the digit input display was available (5 s display), which was then followed by feedback on their performance (2 s display). In this condition, participants were told that the recognition task was of secondary importance. In the memory-only condition, participants were instructed to perform only the recognition task and to ignore the digit input and digit feedback displays. Like previous studies, stimulus images were either close-up or wide-angle

and test views could be either identical or the alternative view. The results of the study

indicated that both conditions elicited memory errors diagnostic of boundary extension, but

that divided-attention produced more pronounced levels (Intraub et al., 2008). Based on these

findings, we hypothesize that when visual attention is allocated exclusively to encoding a

given scene, the resulting representation is high in perceptual weight, allowing the individual

to reject amodal information as perceptually derived with confidence, and thus reducing the

boundary extension error. When visual attention is not entirely devoted to encoding scene

content, however, perceptual information corresponding to a scene would be expected to be

of lower quality, making it more likely that participants would accept amodal information as

perceptually derived. The result of this divided attention is an increased memory error, which

for our purposes manifests in the form of boundary extension; the lower the quality of

perceptual information, the more likely one is to commit this error of commission. This

proposition, that attention allows the viewer to better delineate between sources of

information (e.g., perceptual, amodal), allows one to test the predictions of the multi-source

model of scene perception under conditions of known attentional modulation – in this case,

instances of semantic inconsistency.

In contrast to the overall investment of attention in encoding a scene, it may be that

where attention is allocated during scene perception can affect memory for a given scene. In

a recent ERP study, Martens, Trujillo-Barreto, and Gruber (2011) asked whether attention

may be differentially allocated across a scene across conditions of semantic consistency and

inconsistency. They used a relatively new ERP technique called the steady-state visually

evoked potential (SSVEP) which is an oscillatory response of the visual cortex at the same

temporal frequency (Hz) as the inducing flickering stimulus. Importantly, higher SSVEP

amplitudes are associated with increased attentional allocation to the flickering stimulus (e.g.,

Morgan, Hansen, & Hillyard, 1996; Müller & Hillyard, 2000; Müller Malinowski, &

Hillyard, 2003). By tagging objects and backgrounds with different frequencies (i.e., having

objects and backgrounds presented at 8.6 and 12 Hz, respectively), a Fourier Transform can

reveal differential attentional allocation to objects and backgrounds as indicated by the

relative amplitudes of their corresponding frequencies in the SSVEP. Tagging the objects and

backgrounds with frequencies specifically involves manipulating the luminance of these

portions of the scene such that half of the time it is a lower level of luminance and the other

half of the time it is a higher level of luminance; the frequency that a stimulus is tagged with

determines how many times the luminance is oscillated (e.g., 12 Hz would correspond to 12

oscillations of luminance per second).

In the paradigm used by Martens et al. (2011), participants were simply asked to

detect the presence of a dot that was randomly presented within $4°$ of central fixation, which

was superimposed upon a foreground object presented at 8.6 Hz and a background presented

at 12 Hz. Although the object and background comprising the scenes were irrelevant to the

task, previous studies have provided evidence that attended stimuli superimposed over a

flickering stimulus (in this case a centrally presented dot on top of a scene with a

semantically consistent or inconsistent foreground object) results in higher SSVEP

amplitudes compared to unattended stimuli superimposed over a flickering stimulus, when

distance from fixation is controlled (Morgan et al., 1996; Müller & Hillyard, 2000; Müller et

al., 2003). In both cases, the SSVEP is elicited by the background flickering stimulus, and

not by the superimposed attended vs. unattended stimulus. The findings of the study

(Martens et al., 2011) showed that attention was allocated more to the backgrounds during

consistent object-background presentations, whereas in the inconsistent conditions, attention

was selectively allocated to the foreground object. Based on these findings, we predict that

there will be more boundary extension for inconsistent scenes (Intraub et al., 2008). Similar

to the difference in overall attention allocated to a scene resulting in poorer scene memory

(Intraub et al., 2008), increased levels of attention directed to an inconsistent central

foreground object may result in poorer memory for the background, especially at the

periphery where amodal information can be more readily accepted as perceptually derived. In

other words, increased attention to an inconsistent object may pool attentional resources

away from the background, reducing the amount of perceptual information that can be used

to "defend" against the memory error of boundary extension.

Present Experiments

       The present experiments were intended to address explicitly the possibility of

semantic inconsistency effects upon boundary extension. More specifically, I explored

whether semantic consistency affects boundary extension at short time intervals when scenes

are viewed briefly, and if semantic consistency affects boundary extension at long time

intervals when scenes are viewed for extended periods of time. To address each question in

turn, two experiments were conducted. To explore the effect of semantic consistency on

boundary extension at short time intervals, a paradigm was developed that for each trial, a

single close-up image was presented for 250 ms, masked for 250 ms, and then was followed

by an identical test image with a view-rating task (which will be described in the Method

section). To explore the effect of semantic consistency on boundary extension at longer time

intervals, a paradigm was developed using 10 close-up images that were presented for 15 s

with 1-s masks interleaved. After the presentation of all 10 images, a description of the

memory test was given during the retention interval. The memory test involved the same

rating-task as in the previous experiment, where identical close-up images were presented in

the same sequence as they were initially presented; memory for each picture in the stimulus

sequence was tested. Semantic consistency was manipulated by pairing consistent base

images (e.g., football player in stadium and priest in church) and swapping objects to

produce inconsistent images (e.g., football player in church and priest in stadium). The

rationale for using a 250 ms stimulus and 250 ms mask in the brief duration experiment and

using a 15 s stimulus and 1 s mask in the long-duration experiment was that these are

standard durations used in boundary extension literature, and this allows us to make direct

comparisons to those other studies; it is also standard that about 20 images be presented in

brief-durations test of boundary extension and that 10 images be presented in long-duration

test of boundary extension. The rationale for using color photographs with a single

foreground object was that these types of images may be most sensitive to detecting effects

of semantic consistency, as previous studies have suggested (Bonitz & Gordon, 2008;

Davenport & Potter, 2004). I hypothesized that, for brief-duration pictures and not for longer-

duration pictures, semantically inconsistent scenes would yield larger amounts of boundary

extension compared to semantically consistent scenes because more attention would be

directed to the objects in conditions of semantic inconsistency. I hypothesized that there

would be no significant difference between conditions for longer-duration pictures because

attention would be more evenly spread across the image for 15 s scene presentations. A

significant difference in either direction for long durations may allow this possibility to be

ruled out.

Experiment 1

Method

Participants. Participants consisted of 60 undergraduate students (21 male, 39 female) enrolled in introductory and intermediate-level Psychology courses at Appalachian State University who had volunteered to be a part of the department subject pool to fulfill a research requirement for their course. From this subject pool, participants were recruited using the SONA system, in which students selected from pre-designated time-slots. IRB approval was acquired 10/18/2012 (see Appendix B), and all participants were treated in accordance with ethical guidelines (see Appendix C).

Apparatus. Stimuli were presented using a Dell OPTIPLEX 755 computer and a Dell P-1130 21 in. CRT monitor. The refresh rate was set at 120 Hz, and images were shown at a resolution of 1024×768 pixels in 32 bit color. The program software was based on a template program provided by SR Research Inc., written in C, which was used to display the stimuli and record participants' responses.

Stimuli. Experimental stimuli consisted of 42 images containing a foreground object on a background that were either semantically consistent or inconsistent with the background (e.g., a football player in a football stadium vs. a football player in a church). Twenty of these images were borrowed with permission from Davenport and Potter (2004), and 20 others were composites of objects and backgrounds retrieved from the Internet that were conceptually modeled after other image pairs used in the aforementioned experiment (with one background borrowed from Torralba's indoor scene database; the remaining two practice images were images used in previous boundary extension experiments). The experimental images were borrowed from and modeled after Davenport and Potter (2004) because a

semantic consistency effect has already been found for these images, and we wanted to see if an analogous semantic consistency effect would be found for the spatial expanse of views. Because of the conceptual equivalence between the stimuli, we did not deem it necessary to perform a norming study prior to the present experiments to provide evidence for our categorization of semantically consistent and inconsistent images.

In the original versions of the borrowed images, the object size was different between consistent and inconsistent images (e.g., the priest was larger in the church than on the football field), and because how close-up an image is was operationalized as the size of the object relative to the size of the image view, images were cropped and resized so that objects were the same size in both semantically consistent and inconsistent scenes. After these modifications, image views subtended $11.1° \times 8.3°$ of visual angle and were embedded within a black background which subtended $28.4° \times 21.3°$ of visual angle. Of the 40 experimental stimuli, 20 of the images were semantically consistent (base images) and the other 20 were semantically inconsistent (critical images). Semantic consistency was manipulated in the experimental stimuli by pairing consistent base images (e.g., a football player in a stadium and a priest in a church) and swapping objects to produce semantically inconsistent images (e.g., a football player in a church and a priest in a stadium). Two versions of image sequences were developed such that each participant saw either the consistent or inconsistent versions of image pairs and such that half of the images were inconsistent and the other half were consistent.

Procedure. Participants were seated such that the viewing distance was approximately 80 cm. Before each trial, a fixation cross was presented at the center of the screen at which participants were asked to fixate before the stimulus appears. All 20 trails

(including two practice trials) were self-initiated by a participant pressing the space bar on a keyboard positioned in front of the monitor. On each trial, a single stimulus was presented for 250 ms followed immediately by a 250 ms visual noise mask, which filled the screen. The mask had a dynamically changing central portion that consisted of schematic faces that subtended 5.5° × 5.5° of visual angle. This sequence involved two different faces, with the first shown for 150 ms and the second shown for 100 ms. This type of mask was used to minimize implicit verbalization and to keep the participants' fixation at the center of the image. After the mask, the exact same image reappeared at the same location on the screen, and participants were then prompted to complete a computer-generated rating scale task. In this task, each participant was asked to indicate whether the test view was "the same" view as, "closer up", or "further away" than the stimulus view. These three verbal labels were presented below a horizontal bar at the bottom of the screen (below the test image) that had five equally-spaced intersecting vertical lines that corresponded to the three labels plus the points that bisected each side of the scale.

Participants were instructed to move a vertical pointer along the horizontal bar that could be placed at any point along the horizontal line. The number of pixels between the center of the scale and where the pointer was placed by a given participant was then recorded and normalized to the 5-point scale that corresponds to the five response categories used in the more traditional version of the rating task (i.e., "much closer up" (-2), "somewhat closer up" (-1), "the same" (0), "somewhat farther away" (1), or "much farther away" (2), than the stimulus image). Perceiving the test stimulus as close-up (i.e., negative values) was indicative of boundary extension, whereas boundary restriction was indicated by rating the test stimulus as further away (i.e., positive values). This task was then followed by a confidence rating in

which the participant rated his or her response on the memory test as *sure* (3), *pretty sure* (2),

not sure (1), or *DRP* (0 – "do not remember picture"), and in which, as in the memory test,

only the verbal labels were presented to participants. This task was included to allow

participants to indicate when the completion of the recognition task was based on no memory

of the stimulus image and when it could be assumed that a guessing strategy was used. Trials

on which participants report not remembering the test picture were excluded from all the

analyses. At the end of the experiment, participants were debriefed.

Results

Participants were overall generally confident in their responses (M = 2.06, SD = .36);

there were no cases of participants indicating that they did not remember the picture in the

confidence rating. To test the hypothesis that boundary extension was elicited by the images

used in the experiment, we calculated the average view ratings for all participants collapsed

across both conditions and compared this to a value of zero. A one-sample t-test revealed that

there was significant boundary extension, t(58) = 5.09, p < .001, d = 1.34. This supports the

hypothesis that the images used in this experiment elicited boundary extension. To test the

hypothesis that there would be more boundary extension for inconsistent scenes compared to

consistent scenes, we compared the average view ratings for the two conditions across all

participants. A paired samples t-test revealed that there was no significant difference in

boundary extension between semantically consistent scenes (M = -.17, SD = .28) and

semantically inconsistent scenes (M = -.17, SD = .23), t < 1. These findings do not support

the hypothesis that there would be more boundary extension for inconsistent scenes

compared to consistent scenes.

Experiment 2

Method

Participants. Participants were 60 undergraduates (17 male, 43 female) recruited from the same subject pool as Experiment 1.

Apparatus. The apparatus was the same as in Experiment 1.

Stimuli. The stimuli were the same as in Experiment 1.

Procedure. The viewing distance of participants was the same as in Experiment 1. As opposed to Experiment 1 in which each image was followed by a memory test, in Experiment 2, I administered the memory test after all images had been presented. As in Experiment 1, a fixation cross was initially presented at the center of the screen where participants were asked to fixate before the image sequence began. Participants initiated the sequence of images by pressing the space bar. Images were each presented for 15 s followed by a 1 s visual noise mask that filled the screen. There was a 1-s inter-stimulus-interval with the fixation cross presented as a warning sign for the upcoming onset of the next stimulus. After all of the images had been viewed, the memory task was described to the participants during the retention interval. Participants were asked to complete the same rating scale, followed by a confidence judgment, as in Experiment 1, with the test images presented in the same sequence as presented during encoding. At the end of the experiment, participants were debriefed and thanked for their participation.

Results

Participants were overall generally confident in their responses ($M = 2.26$, $SD = .42$); there were no cases of participants indicating that they did not remember the picture in the confidence rating. To test the hypothesis that boundary extension was elicited by the images

used in the experiment, we calculated the average view ratings for all participants collapsed

across both conditions and compared this to a value of zero. A one-sample *t*-test revealed that

there was significant boundary extension, $t(59) = 5.77$, $p < .001$, $d = 1.50$. This supports the

hypothesis that the images used in this experiment elicited boundary extension. To test the

hypothesis that there is more boundary extension for inconsistent scenes compared to

consistent scenes, we compared the average view ratings for the two conditions across all

participants. A paired samples *t*-test revealed that there was significantly more boundary

extension for semantically consistent scenes ($M = -.33$, $SD = .38$) than for semantically

inconsistent scenes ($M = -.21$, $SD = .34$), $t(59) = -2.10$, $p < .05$, $d = .33$. These findings may

rule out the possibility that there is no difference in boundary extension between semantically

consistent scenes and semantically inconsistent scenes for long durations.

Discussion

In two experiments, I tested whether object-background semantic consistency would

influence boundary extension by presenting scenes with single foreground objects that were

either semantically consistent or inconsistent with the background for brief- or long-duration

image presentations. For brief-duration image presentations, I hypothesized that there would

be more boundary extension for semantically inconsistent scenes because attention would

initially be directed to the object (Bonitz & Gordon, 2008; Martens et al., 2011), resulting in

poorer memory for the spatial expanse of the view (Intraub et al., 2008). For long-duration

scene presentations, I hypothesized that there would be no difference in boundary extension

because attention would be more evenly distributed across the image as time elapsed. Neither

of these hypotheses was supported; here I observed no effect of semantic consistency for

brief durations and instead observed an effect of semantic consistency for longer durations,

with semantically inconsistent scenes eliciting less boundary extension than semantically

consistent scenes.

Research suggests that contextual, semantically related information about a scene can

be rapidly extracted (e.g., Davenport & Potter, 2004), well within the image duration used

here, so why did we find no difference in boundary extension for semantically consistent and

inconsistent object-background relationships at brief image durations and yet a significant

difference at long image durations? One possibility is that boundary extension is context-

independent for brief image durations and is context-dependent for longer image durations.

This interpretation is based on the speculation that rapidly activated, bottom-up influences of

semantic consistency (as posited by early perceptual accounts; e.g., Biederman et al., 1982;

Davenport & Potter, 2004; Friedman, 1979) are not involved in the computation of spatial

expanse, whereas later, top-down influences of semantic consistency (as posited by post-

identification accounts; e.g., Henderson, 1992) are involved in the computation of spatial

expanse. At earlier stages of processing, the computation of spatial expanse may use only

object-background spatial relations, such as how much space an object fills (an idea that will

be discussed in detail later) and may not use any semantic information about the scene. At

later stages of processing, the computation of spatial expanse may involve not only object-

background spatial relations, but also top-down input regarding semantically inconsistent

scenes that could activate two separate contexts (i.e., the context depicted in the scene and

the context in which the inconsistent object normally appears), which together may interfere

with a coherent spatial computation. This interpretation would explain why, in the present

thesis, semantically inconsistent scenes elicited less boundary extension than semantically

consistent scenes for long image-durations only.

Why would boundary extension only be influenced by post-identification input regarding the semantic consistency between objects and backgrounds? The post-identification account (e.g., functional isolation model; Henderson, 1992) posits that semantic information is incorporated into perception only after both objects and their background have been identified. Therefore, this model predicts that there should be no influence of object-background semantic consistency early in perception. Besides the behavioral evidence for this model (e.g., Hollingworth & Henderson, 1998; Henderson, Weeks, and Hollingworth, 1999), the post-identification account of semantic consistency is supported by ERP research reporting a more positive N400 (an ERP component characterized by a negative voltage deflection emerging at approximately 400 ms post-stimulus onset), which is held to reflect semantic violation processing that is post-identification (see Bar, 2004; Ganis & Kutas, 2003; Mudrik, Lamy, & Deouell, 2010). This ERP component has also been associated with the parahippocampal cortex (PHC; see Bar, 2004); the PHC is home to the PPA (see Epstein, 2005) which is thought to be involved in the processing of scenes (Epstein, 2005), and also shows patterns of activity suggestive of its involvement in the computation of spatial expanse (see Park, Brady, Greene, & Oliva, 2011; Park et al., 2007). Given that this later negative potential associated with post-identification semantic violation processing appears to occur in areas thought to be involved in boundary extension (i.e., PPA; Park et al., 2007), a post-identification account of semantic consistency may be appropriate for the computation of spatial expanse; it would follow that early perceptual influences of semantic consistency would not have an effect on boundary extension.

This interpretation would also be consistent with the multiplexer model of contextual

facilitation (Bar, 2004). In this model, coarse global information about a scene is first rapidly

projected to the PHC (Bar, 2004). The PHC uses this coarse information in turn to activate

representations of candidate objects within the inferior temporal cortex (ITC) that belong to

the activated context frame (e.g., office context frame activates object representations of

computer, desk, and chair; Bar, 2004). Subsequently, high-frequency information arrives to

the ITC providing detail about the scene's objects and semantic context. The candidate object

consistent with the detailed information is selected and recognition is completed; all other

candidates are discarded; however, if the object identity projected by the high spatial

frequencies is not one of the candidate objects activated by the PHC, then a N400 is elicited

in the PHC, possibly to suggest an alternative context frame be activated (Bar, 2004). If

spatial computation takes place in this area of the brain, this would lend support to why we

did not find an effect of semantic consistency for brief image durations (i.e., this component

takes place after the image display terminated). That is, spatial computation may operate on

low-spatial frequency information initially and only after approximately 400 ms post-

stimulus onset will a semantic violation be signaled for alternative contexts to be considered

within the spatial framework.

What are the broader implications of these findings for the function of boundary

extension? It has been suggested by Intraub, Hoffman, Wetherhold, and Stoehs (2006) that

boundary extension can serve to integrate successive views based on their finding of more

boundary extension on the cued side of an image compared to the non-cued side. Another

function of boundary extension may be to predict spatial layout to facilitate interaction with

objects, as suggested by the findings of Gottesman (2011), in which distance judgments of

objects were primed by partial views that did not contain the part of the view where the

objects appeared. This claim is also supported by research providing evidence that boundary

extension is not influenced by the absolute size of an object (i.e., whether it's a close-up view

of a small object or a wide-angle view of a large object) but is rather caused by how much

space a given object fills (i.e., spatial relations between object and borders of the depicted

area; see Bertamini, Jones, Spooner, & Hecht, 2005). The findings of Bertamini et al. (2005)

suggest, more specifically, that the closer the distance between the object borders and the

edges of the view, the more boundary extension is elicited (also see Dickinson & Intraub,

2008; Intraub et al., 1992; Intraub et al., 2008; Intraub & Dickinson, 2008; Intraub &

Richardson, 1989; Park et al., 2007). This further suggests that boundary extension may

indeed be functioning to predict spatial layout around an object so that interaction with that

object can be facilitated. If this were the case, then the context in which an object appears

would not influence boundary extension because predicting spatial layout in service of

interaction with a given object should not depend on the semantics of the scene in which a

given object appeared. This implication is consistent with the findings reported here.

Implications for the Multi-Source Model of Scene Perception (Intraub, 2010)

The multi-source model of scene perception (Intraub, 2010) posits that scene

perception is characterized by both visual and non-visual sources of information being

incorporated into a unified framework. Boundary extension under this model is a source-

monitoring error (Johnson et al., 1993) in which non-visual information is falsely attributed

as visually derived. From this, it is expected that the variables that influence boundary

extension should be predicted by the source-monitoring framework (see Johnson et al.,

1993). For example, dividing attention at encoding, which should provide less detailed

perceptual detail and increase reality-monitoring source confusions, has been shown to increase boundary extension (Intraub et al., 2008). As stated in the Introduction, the rationale and hypotheses for this thesis were based on an attention-mediated effect, which made a clear prediction under the source-monitoring framework (Johnson et al., 1993); however, the source-monitoring framework (Johnson et al., 1993) does not make a clear prediction as to how the semantic consistency between objects and backgrounds alone should influence source-monitoring accuracy. This suggests that the source-monitoring framework that is incorporated within the multi-source model of scene perception could be elaborated by including more detail regarding the time course of scene perception and how that would influence source-monitoring performance based on variables such as semantic coherence of a scene.

Alternative Explanations

One alternative explanation of the present findings could be related to when the exact nature of the memory test was revealed. In contrast to Experiment 1, Experiment 2 deferred revealing the exact nature of the memory test until after the encoding procedure; that is, we told the participants that there would be a memory test, but we did not provide the details until all the images had been viewed. Therefore, it may be that participants anticipatively construed the memory task as being one in which they would be asked to indicate which objects and contexts co-occurred with one another. If participants did indeed construe the task in this manner, they would maximize performance by attending more to semantically inconsistent scenes whose objects would not be supported by their probabilistic context at recall. Assuming that this encoding strategy was adopted, the increased attention to semantically inconsistent scenes would explain why we observed less boundary extension for

these conditions in Experiment 2, as the presumably more detailed representations of

semantically inconsistent scenes would reduce memory error for spatial expanse (cf., Intraub,

et al. 2008). Unfortunately, there is no evidence to support or exclude this possibility.

Another alternative would be related to where people fixated during Experiment 2;

because we did not measure eye-movements in Experiment 2, we can only speculate, but it

may be the case that participants made more fixations to the background and the edges of the

view for semantically inconsistent scenes (which should reduce boundary extension;

Gagnier, Dickinson, & Intraub, 2013). Although there is much semantic consistency research

exploring eye movements in multi-object scenes (Bonitz & Gordon, 2008; de Graef et al.,

1990; Henderson et al., 1999; Loftus & Mackworth, 1978), there is little research exploring

eye movements in scenes with a single foreground object. There is evidence arguing against

this possibility, as much research demonstrates that it is the inconsistent objects themselves

that are fixated, and not the inconsistent background in which they appear (e.g., Bonitz &

Gordon, 2008). Moreover, the findings of Martens et al. (2011) suggest that more fixations

should be made to objects in inconsistent scenes than to the backgrounds. Provided that it is

not clear why participants would make more fixations to the background in semantically

inconsistent scenes and that there is evidence to suggest the opposite (more attention to

objects in semantically inconsistent scenes; Martens et al., 2011), I reject this alternative

explanation as a plausible possibility.

Limitations

The first category of limitations in the present experiments is that there are

differences across the two experiments that make isolating the exact cause of the results

difficult. That is, there were many procedural and design aspects that were intentionally not

held constant between the two experiments and that could have individually or collectively

caused the observed differences. For Experiment 1 and Experiment 2, respectively, the

variables confounded with the independent variable of image duration (250 ms vs. 15 s)

included memory load (1 item vs. 10 items), retention interval (250 ms vs. ≈5 min), whether

participants had the chance to actively explore the scene or not (i.e., through eye

movements), and when the memory test was revealed (before encoding vs. after encoding).

The observed effects may have been a result of one or any combination of these variables,

although some are more likely than others. The interpretation developed here is based on the

image duration alone, as it was the manipulated variable; however, the variable of "memory

description" may have caused all or possibly some of this effect as described in the previous

section (also see previous section for rejecting the possibility that participants explored the

image differently across the independent variable in Experiment 2). The variables of memory

load and retention interval may have made access to scene representations in memory more

difficult in Experiment 2; however, it is not clear why there would be better retrieval of

inconsistent scenes relative to consistent scenes.

  The second category of limitations in the present experiments was the properties of

the stimuli used in the present experiments. As mentioned in the Method section, we

borrowed stimuli from Davenport and Potter (2004) because an effect of semantic

consistency had already been detected for this set of stimuli; however, there were two

primary issues with the stimuli that did not make them suitable for testing the possible effect

of semantic consistency on boundary extension in their original state. Both of these issues

stem from the way a scene's spatial expanse has been operationalized in the literature and in

the present thesis. That is, how close-up a scene is was operationalized in this thesis as how

much of the vertical dimension an object filled in the scene; the more of the vertical

dimension an object occupied, the more close-up the scene was. The first issue with the

stimuli with respect to our operationalization of spatial expanse was that objects on average

did not fill up much of the vertical dimension; one of the most replicated findings in the

boundary extension literature is that close-up views elicit more boundary extension compared

to wide-angle views (e.g., Intraub, et al., 1992; Intraub & Richardson, 1989); therefore,

making these images more close-up was desirable because it would provide us with a more

sensitive measure of differences resulting from semantic consistency by maximizing the

amount of boundary extension elicited by the stimuli. Second, some objects were not the

same size when they appeared outside of their original context as they were when they

appeared in their original context (e.g., the priest was smaller in the inconsistent football

stadium scene compared to the consistent church scene); this was an issue because the size

difference of a given object across semantically consistent and inconsistent images would

have likely influenced boundary extension separately from our manipulation of semantic

consistency.

To address these issues, I used Adobe Photoshop to crop the images such that objects

filled more of the vertical dimension so that the overall view was more close-up and such that

objects were the same size in both of the two images that they appeared in. Unfortunately,

controlling for object size through these modifications was at the expense of holding

background spatial expanse constant; that is, semantically consistent and inconsistent scenes

in image pairs sometimes contained different amounts of background spatial expanse (e.g.,

the consistent version of the football field background depicted more of the people in the

bleachers and the top of a house behind the bleachers, whereas the inconsistent version of the

football field background depicted less of the people in the bleachers and did not contain the

top of the house behind the bleachers). This concession may have introduced some error into

our data which may have hidden any effect of semantic consistency on boundary extension

that I was looking for. More specifically, 18 of the total 40 experimental stimuli had

differential amounts of background expanse across semantically consistent and inconsistent

objects; of these 18 images, 16 had more spatial expanse for consistent objects. Given that

40% of the image pairs used here had more spatial expanse depicted for semantically

consistent scenes, this may have significantly reduced the amount of boundary extension for

semantically consistent scenes (less boundary extension for more spatial expanse depicted;

see Gagnier, Intraub, Oliva, & Wolfe, 2011). If this were the case, we would be erroneously

concluding that there was no effect in Experiment 1, when in fact there may have been. For

some of the images, the difference in spatial expanse was so extreme after the cropping that

we abandoned half of the original stimuli and created conceptually identical versions using

images from the Internet. These images had both object size and spatial expanse held

constant across both semantically consistent and inconsistent versions; however, this also

introduced further heterogeneity into our stimuli that may have affected the results in

unknown ways. Finally, the objects in image pairs did not fill the same amount of area (e.g.,

an ambulance filled a more space in a scene than did a camel), which may also have affected

boundary extension independently of our manipulation of semantic consistency.

Suggestions for Future Research

Future research exploring the role of semantic consistency in boundary extension

should control for some of the procedural and stimulus confounds in the present thesis.

Regarding the procedural confounds, experiments that hold all variables constant except for

only one can more confidently isolate what variable, or variables, may be causing the effect of semantic consistency on boundary extension observed in the present thesis. Regarding the stimulus confounds, researchers conducting future research should also create a more homogeneous set of stimuli to test how semantic consistency effects boundary extension. More specifically, this would involve controlling for equivalent amounts of background spatial expanse across semantically consistent and inconsistent scenes, controlling for object area by using roughly equally sized and shaped objects across image pairs used for the semantic consistency manipulation (e.g., a football and a roasted chicken), and creating all scenes using images from the Internet, by cropping all objects and pasting them on other backgrounds, consistent or inconsistent, so that all images contain an object on a different background from which it was photographed (or alternatively using all natural images to avoid the "pasting effect"; Joubert, Fize, Rousselet, & Fabre-Thorpe, 2008). Finally, future research should record eye movements to observe if there are any differences in where people are looking between semantically consistent and inconsistent scenes for long image durations. If future research controls for these procedural and stimulus confounds in addition to measuring eye fixations, we can develop a clearer picture of how semantic consistency influences boundary extension.

Conclusions

Here I have documented the first systematic exploration of the effect of semantic consistency on boundary extension. As a first attempt, the findings reported here are highly suggestive of a memory-based, post-identification influence of semantic consistency on the computation of spatial expanse. This speculative interpretation should be considered with caution, however, as the results were unexpected and have yet to be replicated. Future

research on this topic will provide insight regarding how and when spatial computation is

influenced by semantic information.

References

Bar, M. (2004). Visual objects in context. *Nature Reviews: Neuroscience*, *5*, 617–629.

Bertamini, M., Jones, L. A., Spooner, A., & Hecht, H. (2005). Boundary extension: The role of magnification, object size, context, and binocular information. *Journal of Experimental Psychology: Human Perception and Performance*, *31*, 1288–1307.

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

Biederman, I. (1981). On the semantics of a glance at a scene. In M. Kubovy, & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 213–253). Hillsdale, NJ: Erlbaum.

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

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

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

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

Boyce, S. J., & Pollastek, 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.

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

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

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

Dickinson, C. A., & Intraub, H. (2008). Transsaccadic representation of layout: What is the time course of boundary extension? *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 543–555.

Epstein, R. A. (2005). The cortical basis of visual scene processing. *Visual Cognition*, *12*, 954–978.

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

Gagnier, K. M., Dickinson, C. A., & Intraub, H. (2013). Fixating picture boundaries does not eliminate boundary extension: Implications for scene representation. *The Quarterly Journal of Experimental Psychology*.

Gagnier, K. M., Intraub, H., Oliva, A., & Wolfe, J. M. (2011). Why does vantage point affect boundary extension? *Visual Cognition*, *19*, 234–257.

Ganis, G., & Kutas, M. (2003). An electrophysiological study of scene effects on object identification. *Cognitive Brain Research*, *16*, 123–144.

Gottesman, C. V. (2011). Mental layout extrapolations prime spatial processing of scenes. *Journal of Experimental Psychology: Human Perception and Performance*, *37*, 382–395.

Gottesman, C. V., & Intraub, H. (2002). Surface construal and the mental representation of scenes. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 589–599.

Greene, M. R., & Oliva, A. (2009). The briefest of glances: The time course of natural scene understanding. *Psychological Science*, *20*, 464–472.

Henderson, J. M. (1992). Object identification in context: The visual processing of natural scenes. *Canadian Journal of Psychology*, *46*, 319–341.

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

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

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

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

Hollingworth, A., & Henderson, J. M. (1999). Object identification is isolated from scene semantic constraint: Evidence from object type and token discrimination. *Acta Psychologica*, *102*, 319–343.

Intraub, H. (2010). Rethinking scene perception: A multisource model. *Psychology of Learning and Motivation*, *52*, 231–264.

Intraub, H., Bender, R. S., & Mangels, J. A. (1992). Looking at pictures but remembering scenes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 180–191.

Intraub, H., Daniels, K. K., Horowitz, T. S., & Wolfe, J. M. (2008). Looking at scenes while searching for numbers: Dividing attention multiplies space. *Perception & Psychophysics*, *70*, 1337–1349.

Intraub, H., & Dickinson, C. A. (2008). False memory 1/20th of a second later: What the early onset of boundary extension reveals about perception. *Psychological Science*, *19*, 1007–1014.

Intraub, H., Gottesman, C. V., & Bills, A. J. (1998). Effects of perceiving and imagining scenes on memory for pictures. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 186–201.

Intraub, H., Hoffman, J. E., Wetherhold, C. J., & Stoehs, S. (2006). More than meets the eye: The effect of planned fixations on scene representation. *Perception and Psychophysics*, *5*, 759–769.

Intraub, H., & Richardson, M. (1989). Wide-angle memories of close-up scenes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 179–187.

Irwin, D. E. (1991). Information integration across saccadic eye movements. *Cognitive Psychology*, *23*, 420–456.

Irwin, D. E., & Brockmole, J. R. (2004). Suppressing where but not what: The effect of saccades on dorsal- and ventral- stream visual processing. *Psychological Science*, *15*, 467–473.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*, 3–28.

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

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

Kanizsa, G. (1979). *Organization in vision: Essays on gestalt perception.* New York, NY: Praeger.

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

Mack, M. L., & Palmeri, T. J. (2010). Modeling categorization of scenes containing consistent versus inconsistent objects. *Journal of Vision*, *10*, 1–11.

Martens, U., Trujillo-Barreto, N., & Gruber, T. (2011). Perceiving the tree in the woods: Segregating brain responses to stimuli constituting natural scenes. *The Journal of Neuroscience*, *31*, 17713–17718.

Matin, E. (1974). Saccadic suppression: A review and an analysis. *Psychological Bulletin*, *81*, 899–917.

Morgan, S. T., Hansen, J. C., & Hillyard, S. A. (1996). Selective attention to stimulus location modulates the steady-state visually evoked potential. *Proceeds of the National Academy of Science*, *93*, 4770–4774.

Mudrik, L., Lamy, D., & Deouell, L. Y. (2010). ERP evidence for context congruity effects during simultaneous object-scene processing. *Neuropsychologia*, *48*, 507–517.

Müller, M. M., & Hillyard, S. (2000). Concurrent recording of steady-state and transient event-related potentials as indices of visual-spatial selective attention. *Clinical Neurophysiology*, *111*, 1544–1552.

Müller, M. M., Malinowski, P., Gruber, T., & Hillyard, S. A. (2003). Sustained division of the attentional spotlight. *Nature*, *424*, 309–312.

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

O'Regan, J. K. (1992). Solving the "real" mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology*, *46*, 461–488.

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

Park, S., Brady, T. F., Greene, M. R., & Oliva, A. (2011). Disentangling scene content from spatial boundary: Complementary roles for the parahippocampal place area and lateral occipital complex in representing real-world scenes. *The Journal of Neuroscience*, *31*, 1333–1340.

Park, S., Intraub, H., Yi, D.-J., Widders, D., & Chun, M. M. (2007). Beyond the edges of a view: Boundary extension in human scene-selective visual cortex. *Neuron*, *54*, 335–342.

Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception and Psychophysics*, *16*, 283–290.

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

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

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*, 59–108.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, *62*, 1457–1506.

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.

Silva, M. M., Groeger, J. A., & Bradshaw, M. F. (2006). Attention-memory interactions in scene perception. *Spatial Vision*, *19*, 9–19.

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

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

Yin, C., Kellman, P. J., & Shipley, T. F. (2000). Surface integration influences depth discrimination. *Vision Research*, *40*, 1969–1978.

Appendix A

Table 1

Listing of Stimulus Names

_____

Consistent Version   Inconsistent Version

_____   _____

Intersection – Ambulance   Intersection – Camel
Desert – Camel   Desert – Ambulance
Forest – Deer   Forest – Trumpeter
Parade – Trumpeter   Parade – Deer
Stage – Ballerina   Stage – Cyclist
Road – Cyclist   Road – Ballerina
Football Field – Football Player   Football Field - Priest
Church – Priest   Church – Football Player
Beach – Sand Castle   Beach – Pig
Mud – Pig   Mud – Sand Castle
Race Track – Race Car   Race Track – Tractor
Farm – Tractor   Farm – Race Car
Bowling Alley – Bowler   Bowling Alley – Buffalo
Range – Buffalo   Range – Bowler
Ice Rink – Figure Skater   Ice Rink – Race Horse
Horse Track – Race Horse   Horse Track – Figure Skater
Earth – Space Shuttle   Earth – Sea Turtle
Underwater – Sea Turtle   Underwater – Space Shuttle
Mountain Valley – Woman on Donkey   Mountain Valley – Man on Motorcycle
Parking Lot – Man on Motorcycle   Parking Lot – Woman on Donkey

_____

Note. Images in pairs follow each other in the rows.

Appendix B

To: Daniel Lacombe

CAMPUS MAIL

From: Dr. Stan Aeschleman, Institutional Review Board Chairperson
Date: 10/18/2012
RE: Notice of IRB Approval by Expedited Review (under 45 CFR 46.110)
Study #: 13-0086

Study Title: Semantic consistency in scene perception
Submission Type: Initial
Expedited Category: (7) Research on Group Characteristics or Behavior, or Surveys, Interviews, etc.
Approval Date: 10/18/2012
Expiration Date of Approval: 10/17/2013

This submission has been approved by the Institutional Review Board for the period indicated. It has been determined that the risk involved in this research is no more than minimal.

Investigator's Responsibilities:

Federal regulations require that all research be reviewed at least annually. It is the Principal Investigator's responsibility to request renewal of approval before the expiration date. You may not continue any research activity beyond the expiration date without IRB approval.

Any adverse event or unanticipated problem involving risks to subjects must be reported immediately to the IRB. You are required to obtain IRB approval for changes to any aspect of this study before they can be implemented except to eliminate apparent immediate hazards. Best wishes with your research!

CC:
Christopher Dickinson, Psychology

Appendix C

Consent to Participate in Research
Information to Consider About this Research

Semantic Consistency in Scene Perception

Principal Investigator: Daniel LaCombe Jr.
Department: Psychology
Contact Information: 311 Smith-Wright Hall, lacombed@appstate.edu
Faculty advisor: Dr. Chris Dickinson
Contact Information: 203 Smith-Wright Hall, 828 263-4010, dickinsonca@appstate.edu

What is the purpose of this research?

You are being invited to take part in a research study about the nature of information that is retained when we view pictures that depict scenes in the real world. By doing this study we hope to learn about what information people remember when viewing pictures of natural scenes.

Why am I being invited to take part in this research?

You are being invited to participate because you are a healthy volunteer with no history of vision or attention deficits. If you volunteer to take part in this study, you will be one of about 240 people to do so.

Are there reasons I should not take part in this research?

If you have any uncorrected vision problems, problems with color vision, or attention deficits that might affect performance in this experiment, you will inform the experimenter that you may not be eligible to participate. You are not required to disclose the actual reason, however.

What will I be asked to do?

The research procedures will be conducted at 216 Smith-Wright Hall. You will need to come here one time for approximately 30 minutes during the study. You will be asked to view a series of single photographs of natural scenes, with each picture followed by a mask and a brief memory test for that picture.

What are possible harms or discomforts that I might experience during the research?

To the best of our knowledge, the risk of harm for participating in this research study is no more than you would experience in everyday life.

What are possible benefits of this research?

There may be no personal benefit from your participation but the information gained by doing this research may help others in the future.

This study should help us learn about what information about scenes is remembered across eye movements. In addition, your participation may contribute to overall knowledge about how people study and remember scenes.

Will I be paid for taking part in the research?

We will not pay you for the time you volunteer while being in this study. You will receive 1 Experiential Learning Credit (ELC) toward your General Psychology research participation requirement for today's experiment (if you are participating for credit in another class, you will receive 1 ELC for that class). The requirements and options for research participation have been outlined in the syllabus for your psychology class.

How will you keep my private information confidential?

Your information will be combined with information from other people taking part in the study. When we write up the study to share it with other researchers, we will write about the combined information. You will not be identified in any published or presented materials. This study is anonymous. That means that no one, not even members of the research team, will know that the information you gave came from you. No identifying information will be part of the data file from today's experiment, and the list that identifies you as a participant in this experiment will be kept in a password-protected database to help maintain your confidentiality. Once the information in our database that identifies you is no longer needed, the list containing that information will be destroyed.

Whom can I contact if I have a question?

The people conducting this study will be available to answer any questions concerning this research, now or in the future. You may contact the Principal Investigator at lacombed@appstate.edu. If you have questions about your rights as someone taking part in research, contact the Appalachian Institutional Review Board Administrator at 828-262-2130 (days), through email at irb@appstate.edu or at Appalachian State University, Office of Research and Sponsored Programs, IRB Administrator, Boone, NC 28608.

Do I have to participate? What else should I know?

Your participation in this research is completely voluntary. If you choose not to volunteer, there will be no penalty and you will not lose any benefits or rights you would normally have. If you decide to take part in the study you still have the right to decide at any time that you no longer want to participate. There will be no penalty and no loss of benefits or rights if you decide at any time to stop participating in the study. This research project has been approved, as required, by the Institutional Review Board of Appalachian State University.

This study was approved on 10/1812. This approval will expire on 10/17/13 unless the IRB renews the approval of this research.

I have decided I want to take part in this research. What should I do now?

By proceeding with the activities described above, you acknowledge that you have read and agreed to the descriptions and terms outlined in this consent form, and voluntarily agree to participate in this research.

Vita

Daniel Charles LaCombe Jr. was born in Newark, Delaware, to Andea and Dan LaCombe. He graduated from the University of Delaware in 2011 with a Bachelor of Arts degree in Psychology. Immediately afterwards, he attended Appalachian State University where he was awarded a Master of Arts degree in General Experimental Psychology.