Neural Representations of the Visual World

Across Space and Time

By

Matthew X. Lowe

A thesis submitted in conformity with the requirements
for the degree of Doctorate of Philosophy

Department of Psychology
University of Toronto

© Copyright by Matthew X. Lowe 2018
Neural Representations of the Visual World Across Space and Time

Matthew X. Lowe

Doctorate of Philosophy

Department of Psychology
University of Toronto

2018

Abstract

A fundamental question of cognition and perception has asked how the human brain represents the local visual environment. Investigations seeking to answer this question have often explored the contributions of spatial information, such as geometry and layout, in forming a holistic percept of the visual world, yet a growing body of work suggests multiple properties may be crucial to the underlying neural mechanisms supporting scene recognition. This dissertation aims to examine the role of common and ubiquitous visual features, described here as geometry and surface properties, in shaping neural representations of the visual environment across space and time. Investigations in Chapter 2 sought to elucidate the weighting of geometry and surface properties in object and scene perception, and demonstrated equal sensitivity in the scene-selective parahippocampal place area for these features within a scene, and greater sensitivity to the surface information of an object, over its geometry. In Chapter 3, the interactions of these properties with goal-directed behaviour and higher-order scene attributes were examined, with results indicating that scene-selective cortex constructs a flexible representation of the visual environment by integrating diagnostic visual information with task context. Following this investigation, Chapter 4 examined these features within the temporal domain, wherein dissociable neural activity patterns emerged demonstrating how these features may diverge
across time to influence the time course of scene perception. Together, these results demonstrate that multiple visual properties shape neural representations responsible for successful scene recognition, and these properties interact with goal-directed behaviour, and unfold across time, to influence our perception of the visual world.
Acknowledgements

Since joining his lab, my primary advisor, mentor, and friend, Jonathan S. Cant, has been an incredible source of support and guidance, both academically and personally. I am immensely grateful to have had Jonathan as a mentor, and feel privileged to have worked with him these past years. My co-supervisor, Susanne Ferber, whom I have been fortunate to have worked with prior to and throughout my graduate studies, has supported and encouraged growth on both a professional and personal level, and I am incredibly thankful for her wisdom and guidance. These words could not begin to describe how grateful I am to have had the experiences provided by Susanne and Jonathan. Together, they have afforded boundless opportunities, challenged me to succeed, and inspired and fostered my love of research.

To my collaborators and fellow lab members, I am indebted to you for making my graduate studies such an enjoyable experience. Thank you to Kristin E. Wilson and Justin Ruppel, for welcoming me many years ago, and for the constant conversation, humour, and support since. Thank you to Dirk B. Walther, Jason Rajsic, Jason P. Gallivan, Morris Moscovitch, Ryan A. Stevenson, Morgan D. Barense, Jessica Robin, and the numerous research assistants who have contributed to these projects.

Most of all, I am fortunate to have had incredible support and love from my family, especially my mother, Lyndsay Jackson, and my sister, Sarah Lowe. Lyndsay has challenged and pushed me to succeed from the beginning, and it is difficult to express the profound happiness I have had in sharing this journey with her, which was only possible through her support. You mean the world to me. Thank you.
# Table of Contents

Abstract ........................................................................................................................................... ii

Acknowledgments ............................................................................................................................... iv

Table of Contents ................................................................................................................................. v

List of Tables .......................................................................................................................................... vi

List of Figures ......................................................................................................................................... vii

List of Appendices ................................................................................................................................. x

Chapter 1: Introduction .......................................................................................................................... 1

Chapter 2: Neural Representation of Geometry and Surface Properties in Object and Scene Perception.. 11
  Method .................................................................................................................................................. 12
  Experiment 1: Feature Processing in Objects and Scenes Separately ................................................. 19
    Results and Discussion ...................................................................................................................... 21
  Experiment 2: Feature Processing in Scenes Containing an Object ................................................... 25
    Results and Discussion ...................................................................................................................... 26
  Experiment 3: fMR-adaption of Scene Features .................................................................................. 29
    Results and Discussion ...................................................................................................................... 31

Chapter 3: Feature Diagnosticity and Task Context Shape Activity in Human Scene-selective Cortex.... 38
  Method .................................................................................................................................................. 40
  Results and Discussion ...................................................................................................................... 45

Chapter 4: Temporal Dynamics Underlying Scene Categorization in the Human Brain .................... 58
  Method .................................................................................................................................................. 59
  Results and Discussion ...................................................................................................................... 63

Chapter 5: General Discussion ............................................................................................................. 76

References .............................................................................................................................................. 92

Appendices ........................................................................................................................................... 103
List of Tables

Table 1 – Accuracy (percent correct) and reaction time (ms) for each condition in each Experiment (1-3) .......................................................... 35
Table 2 – Accuracy (percent correct) for each condition .......................................................... 56
Table 3 – Recognition performance for each category of photographs and line drawings............ 74
List of Figures

Figure 1 – (A) Examples of the stimuli used in Experiment 1. Scenes and objects are defined by their shape (circular vs. square) and texture (wood vs. brick). Observers attended to the shape or texture of the object or scene, either of which could change while the other was held constant. (B) Examples of stimuli used in Experiment 2. The stimuli and procedure were identical to Experiment 1, with the exception that objects were placed within scenes. (C) Examples of the stimuli used in Experiment 3. Scenes could vary across 10 different shapes, and 10 different textures. In Experiment 3, observers attended only to overall changes across images, and did not attend directly to any one particular feature. For additional examples of the stimuli used in Experiment 3, see Appendix Item 1. 

Figure 2 – Functionally defined ROIs are shown on the brains of two representative observers. Talairach coordinates for peak voxels of each ROI in observer one (PPA – FFA) and two (OFA) are as follows: LPPA, -29, -34, -10; RPPA, 27, -34, -5; LRSC, -15, -56, 14; RRSC, 14, -59, 17; LOPA, -41, -72, 14; ROFA, 32, -80, 0; RLO, 39, -71, 5; LFFA, -43, -36, -17; RFFA, 41, -47, -13; LOFA, -36, -79, 4; ROFA, 43, -75, -3.

Figure 3 – Univariate response amplitudes for Experiment 1. BOLD signal activation for all conditions (attend scene shape; attend scene texture; attend object shape; attend object texture) in each ROI. Data are represented as mean ± SEM.

Figure 4 – (Above) Confusion matrices for Experiment 1. (A) Confusion matrices (chance = 0.25) for Experiment 1 generated from multiclass discriminations showing the distribution of classification errors across all four conditions (SS = Scene Shape; ST = Scene Texture; OS = Object Shape; OT = Object Texture) for each ROI. The average classifier response proportions across observers are shown. When decoding is perfect, the confusion matrix will have a diagonal containing values of 1 and the rest of the matrix will be 0. Note that the average decoding performance (shown in Fig. 4B) is defined as the mean across the diagonal. To highlight differences in decoder performance, the matrices have been rescaled (rather than being scaled from 0-1) (Below) Multivariate results for Experiment 1. (B) Overall classification accuracy (chance = 25%; dashed line) of all four conditions for each ROI. (C) Decoding accuracy (chance = 50%; dashed line) for scene feature discriminations (shape vs. texture), object feature discriminations (shape vs. texture), shape category discriminations (scene vs. object) and texture category discriminations (scene vs. object) for each ROI. Data are represented as mean ± SEM. *p < 0.05, **p < 0.01, ***p < 0.001, all p-values shown have been FDR-corrected.

Figure 5 – Univariate response amplitudes for Experiment 2. BOLD signal activation for all conditions (attend scene shape; attend scene texture; attend object shape; attend object texture) in each ROI. Data are represented as mean ± SEM.

Figure 6 – (Above) Confusion matrices for Experiment 2. (A) Confusion matrices (chance = 0.25) for Experiment 2 generated from multiclass discriminations showing the distribution of classification errors across all four conditions (SS = Scene Shape; ST = Scene Texture; OS = Object Shape; OT = Object Texture) for each ROI. The average classifier response proportions across observers are shown. When decoding is perfect, the confusion matrix will have a diagonal containing values of 1 and the rest of the matrix will be 0. Note that the average decoding performance (shown in Fig. 6C) is defined as the mean across the diagonal. To highlight differences in decoder performance, the matrices have been rescaled (rather than being scaled from 0-1). (Below) Multivariate results for Experiment 2. (B)
Classification accuracy (chance = 25%; dashed line) of all four conditions for each ROI. (C) Decoding accuracy (chance = 50%; dashed line) for scene feature discriminations (shape vs. texture), object feature discriminations (shape vs. texture), shape category discriminations (scene vs. object) and texture category discriminations (scene vs. object) for each ROI. Data are represented as mean ± SEM. *p < 0.05, **p < 0.01, ***p < 0.001, all p-values shown have been FDR-corrected.

Figure 7 – fMR-Adaptation response amplitudes for Experiment 3. Levels of adaptation across conditions (no change; same-shape-different-texture; different-shape-same-texture; both change) for each ROI. Data are represented as mean ± SEM.

Figure 8 – Experimental Stimuli. (A) Examples of the four scene categories used. Scenes were defined by their spatial boundary (open vs. closed) and content (natural vs. manufactured). Observers attended to either the global texture or spatial layout (task context) within a scene, either of which could change while the other was held constant. (B) Comparisons of structural information displayed through line drawings. Note that, in the absence of textural information, structure is more informative in manufactured scenes.

Figure 9 – Regions of Interest. Functionally-defined ROIs are shown on a representative observer’s brain. Talairach coordinates for peak voxels of each ROI in this representative observer are shown as follows: LPPA, -22, -43, -9; RPPA, 31, -45, -5; LRSC, -16, -54, 7; RRSC, 22, -49, 7; LOPA, -29, -84, 15; ROPA, 40, -83, 17; LLO, -39, -72, -7, RLO: 45, -69, 2; LFFA, -36, -34, -23; RFFA, 45, -45, -12; LEVC, -6, -87, -3; REVC, 10, -84, -4.

Figure 10 – Univariate Results. BOLD signal activation for natural (N) and manufactured (M) scenes when attending to either layout or texture. **p = <.01, ***p = <.001.

Figure 11 – Multivariate Results. (A) Classification accuracy (chance = 12.5%; dashed line) of all eight conditions for each ROI. (B) Decoding accuracy (chance = 50%) of scene content with spatial boundary held constant when attending to layout, and (C) spatial boundary with scene content held constant when attending to layout. (D) Decoding accuracy of scene content with spatial boundary held constant when attending to texture, and (E) spatial boundary with scene content held constant when attending to texture. Error bars represent the standard error of the mean. *p = <.05, **p = <.01, ***p = <.001, +q ≤ .05.

Figure 12 – Confusion matrices generated from the multiclass discriminations (A) Hypothetical confusion errors within the same Spatial Boundary or within the same Content. Note that when decoding is perfect, the confusion matrix will have a diagonal containing values of 1 and the rest of the matrix will be zero. (B) Confusion matrices for each ROI representing classifier errors across conditions. The average classifier response proportions across observers are shown. Shaded squares represent predicted decoding values significantly greater than 12.5% (chance). (C) Confusion errors (collapsed across task) within the same spatial boundary or within the same content across ROIs. Error bars represent the standard error of the mean. N = Natural, M = Manufactured, O = Open, C = Closed.

Figure 13 – Stimuli examples from each of the six scene categories as colour photographs (above), and their corresponding line drawings (below).

Figure 14 – Site selection (above) was based on overall higher right-hemisphere response amplitudes using scalp maps averaged across observers for colour photographs and line drawings (below) from 100-800ms. Sites included a centro-occipital site (Oz), a right centro-parietal site (P4), and a right centro-frontal site (F4), indicated in red.
Figure 15 – Event-related potentials (ERPs) and their difference waves plotted in microvolts (μV) for natural (beach, forest, mountain) versus manmade (city, highway, office) scenes averaged across observers (N = 16) for centro-occipital (Oz), right centro-parietal (P4), and right centro-frontal (F4) sites. An asterisk indicates the earliest significant time point (p < 0.01) within a time component................................................................. 65

Figure 16 – Event-related potentials (ERPs) and their difference waves plotted in microvolts (μV) for open (beach, highway) versus closed (forest, city) scenes averaged across observers (N = 16) for centro-occipital (Oz), right centro-parietal (P4), and right centro-frontal (F4) sites. An asterisk indicates the earliest significant time point (p < 0.01) within a time component................................................................. 66

Figure 17 – Event-related potentials (ERPs) plotted in microvolts (μV) for all scene categories (beach, forest, mountain, city, highway, office) averaged across observers (N = 16) for centro-occipital (Oz), right centro-parietal (P4), and right centro-frontal (F4) sites................................. 68

Figure 18 – Similarity matrices showing the grand-averaged correlations of each scene category for all observers (N = 16) averaged across time (0 – 1000 ms) in centro-occipital (Oz), right centro-parietal (P4) and right centro-frontal (F4) electrode sites. (B) Independent samples t-tests comparing diagonal and off-diagonal values. In each of these cases, r values were transformed to a normal distribution using the inverse hyperbolic tangent (Fisher’s z transform). *p < 0.05, **p < 0.01, ***p < 0.001.................................................. 69

Figure 19 – Grand-averaged discriminations (N = 16) of scene categories (diagonal vs. off-diagonal) for colour photographs (left) and line drawings (right) and their difference waves in centro-occipital (Oz), right centro-parietal (P4) and right centro-frontal (F4) electrode sites. Asterisks (marked in red) indicate significant discriminations (Fisher Z) across scene categories using a moving-window containing 30 samples from ERPs in each window (spanning 58.6ms). Dashed lines represent the earliest time of significant category discrimination at any of the sites. *p < 0.01 ................................................................. 71

Figure 20 - Similarity matrices showing the grand-averaged correlations of each scene category for all observers (N = 16) averaged across time (0 – 1000 ms) in centro-occipital (Oz), right centro-parietal (P4) and right centro-frontal (F4) electrode sites in both colour photographs and line drawings................................. 72

Figure 21 - Grand-averaged discriminations (N = 16) of scene categories (diagonal vs. off-diagonal) comparing colour photographs and line drawings and their difference waves in centro-occipital (Oz) region. Asterisks (marked in red) indicate significant discriminations (Fisher Z) across colour photographs and line drawings using a moving-window containing 30 samples from ERPs in each window (spanning 58.6ms). *p < 0.01................................. 73
Appendix Item 1 – Additional Examples of Experimental Stimuli. An example of each type of geometry (room shape) and each type of surface property (wall texture) used in Experiment 3. Each room was displayed with all surface properties for a total of 100 unique stimuli ........................................................................................................................................ 103

Appendix Item 2 - ERPs (N = 16) with 95% confidence intervals plotted for scene content (natural; manmade) with significant temporal clusters (asterisks) highlighted, indicating significant differences between natural and manmade scenes, *p < 0.01........ 104

Appendix Item 3 - ERPs (N = 16) with 95% confidence intervals plotted for scene boundary (open; closed) with significant temporal clusters (asterisks) highlighted, indicating significant differences between open and closed scenes, *p < 0.01 ................. 104
Chapter 1
Introduction

One of the most astonishing aspects of human perception is the ability to efficiently perceive and understand the visual world, from dense forests to soaring cityscapes, in only a single glance. Research exploring this question has revealed that humans have the ability to extract the meaning, or gist, from an image presented for as little as 13ms (Potter, Wyble, Hagmann and McCourt, 2014). The remarkable speed at which this process occurs is consistent with a feedforward model of visual perception, in which an initial wave of neural activity captures much of the information needed to extract meaning and allow for identification of our environment (VanRullen and Thorpe, 2001). This feedforward sweep is followed by a slower and more iterative feedback process which may further shape our understanding of the visual world. A fundamental question of visual perception has therefore been how the brain represents scene information to perform this feat, and which information is drawn from a broad array of cues to construct a stable and efficient representation of the surrounding environment which can be deduced within a fraction of a second. These questions have been explored for decades using behavioural methods, but since the initial description of the scene-selective parahippocampal place area (PPA; Epstein and Kanwisher, 1998), investigations have sought to answer this question by attempting to clarify the nature of the neural representations in this region, and other scene-selective regions within the ventral visual cortex. To explore how neural representations of the local visual environment are shaped, this dissertation will examine how different scene features are represented within the PPA, in addition to other areas of scene- and object-selective ventral visual cortex. Following these investigations, this dissertation will examine scene processing within the temporal domain, to investigate the time course of visual information processing and scene categorization.

Scene-selective areas within the ventral visual cortex

In 1998, an early investigation of the human parahippocampal cortex reported a region responding selectively and automatically to passively viewed scenes, such as landscapes and cityscapes, over non-scene single objects or faces (Epstein and Kanwisher, 1998). This region, named the parahippocampal place area (PPA), responded with greater strength to intact over
scrambled or fractures scenes, and it was therefore suggested that this region was particularly sensitive to the global spatial information, or layout, within a scene. Emerging with the initial description of the PPA, the influential spatial layout hypothesis specifies this region represents the geometric structure of a scene as defined by its background elements, and therefore eschews surface information such as the colour and texture of a scene. Since this description, much of the research investigating this region has revealed a primary role for the PPA in the encoding of spatial features within a scene, such as structural geometry or layout (Epstein and Kanwisher, 1998; Epstein et al., 2003), spatial boundary (Park et al., 2011), and spatial depth (Kravitz et al., 2011). These studies have supported the idea that processing within the PPA is driven by spatial information, and this information may be fundamental to rapid scene understanding.

Recent evidence, however, has begun to support the notion that the neural representations in the PPA extend beyond spatial features and include the encoding of surface texture and material properties (Peuskens et al., 2004; Cant and Goodale, 2007; 2011). This region has been further shown to represent non-spatial contextual associations of objects (Aminoff et al., 2007, Bar et al., 2008) and events (Diana, 2016), and high-level conceptual scene categories (Walther et al., 2009; 2011; Dilks et al., 2011). Scene-selective cortex may therefore represent multiple spatial and nonspatial visual properties, such as geometry and surface information, which in turn jointly support higher-level scene understanding, such as conceptual categories and affordance-related information. Disentangling and directly investigating the unique contributions of individual visual elements to scene representation and understanding has been a central challenge in visual cognition, however, and previous research has yet to elucidate how different visual features shape underlying neural responses both within the spatial realm, and within the temporal domain.

Since the initial description of the PPA, other areas of scene-selective cortex have been identified within the ventral visual cortex. Over a century ago, Brodmann (1909) delineated the human retrosplenial cortex within the posterior cingulate cortex. The retrosplenial complex (RSC), a functionally-defined scene-selective region which is not necessarily identical to the anatomically-defined retrosplenial cortex, was later found to show strong activity during scene viewing and scene imagery (O’Craven and Kanwisher, 2000; Bar and Aminoff, 2003). In contrast to the PPA, RSC shows greater sensitivity to familiar than to unfamiliar places (Epstein et al., 2007; Suguira et al., 2005), therefore suggesting this region is involved in long-term spatial
knowledge of familiar environments (Epstein, 2008). The RSC has been linked most prominently to spatial navigation, as navigational difficulties have been frequently reported when the RSC is damaged (e.g., Takahashi et al., 1997; Katayama et al., 1999; Ino et al., 2007). Specifically, the RSC may represent internal spatial representations such as location and direction, affording humans the ability to orient to the local visual environment during navigation (Marchette et al., 2014). In addition to spatial cognition, this region may also play a key role in contextual and episodic memory, suggesting the RSC supports the processing of both sensory-perceptual information and consolidation for long-term memory (see Miller et al., 2014).

Most recently, an additional scene-selective region has been identified near the transverse occipital sulcus (TOS) (Grill-Spector, 2003). This functionally-defined region, named the occipital place area (OPA), has since been causally linked to scene recognition and identification (Dilks, Julian, Paunov, and Kanwisher, 2013). Whereas its precise role in scene perception is presently unclear, recent investigations have demonstrated the causal involvement of the OPA in boundary perception (Julian et al., 2016), and have suggested that this region may be involved in representing the local elements of a scene, and in first-person perspective for visually-guided navigation (Kamps et al., 2016a; 2016b). The present dissertation will focus on the PPA, but will additionally explore these other scene-selective regions to comprehensively investigate visual scene processing and its underlying neural mechanisms.

A role for surface properties in scene recognition

Akin to structural features, surface properties are ubiquitous within a scene, and inform our general perception and recognition of the world around us. Evidence from neuropsychology has supported this claim, highlighting a role for surface information in fundamental aspects of recognition. For instance, previous research has indicated that a patient with profound visual form agnosia, which manifested as an impairment in the ability to process visual structure, could use visual texture and colour information for accurate scene recognition, suggesting these visual features play an important role in the formation of scene identity (Steeves et al., 2004). In object perception, while it has been argued that edge-detection is primary to surface perception, which may play only a secondary role in perception, visual texture may be especially important in defining edge and contour information used for finding partially occluded objects in complex and crowded environments (Biederman and Ju, 1988). Surface characteristics such as texture
may therefore facilitate visual search by defining edge information necessary for the extraction of meaning, and may interact with edge information to form a stable representation of the visual world.

Texture is also instrumental in providing visual cues which aid in identification and action planning necessary for interacting with objects in our environment (Buckingham et al., 2009; Gallivan et al., 2014), such as providing cues to their weight and other material properties, and may even form a contextual bridge linking objects and their surrounding scenes within the same context (Lowe, Ferber, and Cant, 2015). For example, the local geometry of objects (i.e. shape) may provide cues which impair or facilitate the processing of global geometry (i.e., layout), while the processing of surface properties may have a global-advantage which may influence the processing of local surface information. While spatial information is an important aspect of visual recognition, mounting evidence suggests that it is not the only important cue for scene understanding. For instance, a growing body of work suggests surface characteristics such as colour and texture are instrumental in mediating early-stage scene gist processing responsible for successful scene recognition (Schyns and Oliva, 1994; Oliva and Schyns, 2000; Goffaux et al., 2005; Castelhano and Henderson, 2008), anticipating a role for surface properties in shaping underlying neural representations of the local visual environment. In fact, investigations have shown that areas overlapping with the scene-selective PPA are involved in processing the surface texture and material properties of isolated objects (Peuskens et al., 2004; Cant and Goodale, 2007; 2011). Although numerous behavioural studies have indicated an important role for surface information in shaping scene recognition and understanding, there have been fewer neural investigations which have explored the role of surface information in scene recognition directly.

**A framework for scene understanding**

One framework for scene perception which may reconcile these differences proposes that the recognition of complex visual scenes can be understood through interactions between perceptually available information and categorization demands (Oliva and Schyns, 1997). This recognition framework centers on the notion of feature diagnosticity: the idea that specific visual cues are used for specific types of categorizations, and an interaction between task demands and
available visual information can explain how different cues are used to recognize scenes. In other words, diagnostic visual features may emerge as a function of their usefulness in defining the identity of a scene, and the task demands placed on the observer. Task-dependent demands based on observer goals, such as search, detection, recognition, and action, may directly influence perception and the utilization of specific visual cues within a scene. Given the variability of visual information across scene categories (e.g., beaches, forests, city streets, etc.), Oliva and Torralba (2006) proposed that the most effective global features for scene identification will be those capturing the global structure and meaning of the visual world. For example, modern cities are dominated by prominent edge-based information from structures containing straight horizontal and vertical lines, yet natural landscapes have zones of characteristic textures and undulating contours which may be meaningful for scene identification (Oliva and Torralba, 2001). Spatial layout may thus be especially important for perception and navigation within a modern city containing meaningful rectangular geometric arrangements which constitute buildings and streets, yet a forest can be described in terms of the roughness and homogeneity of its textural components, providing meaningful information to a human observer comparing natural scenes (Rao and Lohse, 1993). Since we are not merely passive observers within our environment, behavioural intents and goals may have an important influence in how we perceive and interact with the world around us. These goals may be flexible and dynamic according to current circumstances and surroundings, and may therefore influence the underlying neural mechanisms of perception.

Scene processing within the temporal domain

Despite behavioural evidence highlighting the importance of both geometry and surface information for scene recognition, there is some neural evidence to suggest that structural features may precede the processing of surface information and may be the primary driving force behind our understanding of the visual world. Recent investigations have revealed that in addition to early visual cortex, the PPA is part of a network of regions affording the ability to distinguish among natural scene categories, such as a forest or city (Walther et al., 2009). Critically, neural categorization accuracy for these scenes was found to be similar between colour photographs and line drawings (Walther et al., 2011), suggesting that the structure
preserved in line drawings is sufficient for scene categorization even in the absence of surface information such as colour and texture. To fully understand the mechanisms underlying these processes, however, we must not only seek to explore them across space, but also across time. Within the spatial domain, these features may be closely interwoven in the human brain and processed within shared or overlapping neural regions, affording us limited precision when exploring their unique neural markers. Within the temporal domain, however, they may diverge along a hierarchy of visual processing and unfold across time for the purposes of scene recognition. Yet these possibilities necessitate further examination to obtain a greater understanding of the neural mechanisms underlying scene recognition.

Fewer investigations have explored scene recognition and categorization within the temporal domain, yet there have been considerable strides made seeking to understand the temporal dynamics of visual perception. Pioneering work using electroencephalography (EEG) has shown event-related potentials (ERPs) reflect properties of a visual stimulus, such as identity and category, within the first 150ms following the presentation of an image (Thorpe, Fize, and Marlot, 1996; Vanrullen and Thorpe, 2001). Similarly, intracranial electrophysiological recordings have reported visual category selectivity within only 100ms in both monkeys (Vogels, 1999) and humans (Liu et al., 2009), providing evidence that meaning may be extracted rapidly within the visual processing stream. Within scene perception, evidence from EEG and magnetoencephalography (MEG) in humans has revealed sensitivity to global scene properties, such as spatial boundary, or layout (i.e., whether a scene contains open boundaries, such as a desert, or closed boundaries, such as a forest), and content (i.e., whether a scene is natural, such as a desert, or manmade, such as a city), within ~250ms from stimulus onset, and single-image discrimination within only 100ms (Groen et al., 2013; Cichy et al, 2016; Harel et al., 2016). The time course for basic-level categorization (i.e., whether a scene is described as a beach or a city) of scenes, however, may differ from these higher-order, or global descriptions.

There is some behavioural research which suggests that basic-level categorization occurs prior to global distinctions and may therefore form a fundamental basis for how we describe our environment (Rosch et al., 1976; Tversky and Hemenway, 1983), yet there is also work which counters this claim and states the opposite is true (Oliva and Torralba, 2001; Green and Oliva, 2009; Loschky and Larson, 2010; Kadar and Ben-Shahar, 2012). In contrast, previous
behavioural evidence has suggested that different levels of category distinctions require the same information and processing speed for recognition (Fei-Fei, Koch, and Perona, 2007), and that a basic or global category advantage may be flexible and depend on the scene category observed (Banno and Saiki, 2015). There is therefore much debate surrounding the temporal dynamics of scene categorization in the human brain. It may be likely that different visual properties influence the temporal dynamics of scene recognition for both basic-level (e.g., beach; forest) and global (e.g., boundary; content) scene categorization. For example, there is some evidence to suggest that edge-based information may receive priority-processing over surface information within the visual stream during scene recognition, and may therefore play a crucial role in categorization (Fu et al., 2016). The precise temporal dynamics of this relationship, however, are unclear, presenting a challenge to our understanding of how categorization of the visual world around us unfolds across time.

**Thesis overview**

To address the role of visual feature information in shaping neural representations of the visual world across space and time, this dissertation will present a series of experiments which aim to understand the role of geometry and surface properties in shaping underlying neural representations in high-level visual cortex. These questions seek to explore how these visual features interact with global scene properties to form a scene percept within the brain, and how diagnosticity and task demands may influence these neural mechanisms. In addition to examining the roles of geometry and surface properties in shaping representations of the visual world within the spatial domain, this research also seeks to understand how these features interact across the temporal domain to shape scene recognition and categorization across time. To explore these topics, the studies presented in this dissertation seek to address three broad questions.

First, how does visual feature information contribute to the neural processes underlying visual scene perception? To address this question, I will examine the role of geometry and surface properties in shaping the underlying neural representations of scenes, and the interaction between object and scene information (Lowe et al., 2015), within the ventral visual cortex (Chapter 2). To accomplish this, a novel set of images specifically designed to explore the
relative weighting (i.e., levels of univariate activation) of geometry and surface properties in object and scene perception will be used to compare neural representations of these features across objects and scenes. This investigation will first test the hypothesis that the PPA will show equal weighting (i.e., equivalent levels of activation) to the processing of the geometry and surface properties of a scene when scenes and objects are presented in isolation, but greater sensitivity to the surface properties of an object over its shape, replicating and extending previous findings (Cant and Goodale, 2007; 2011).

Building on previous behavioural research which suggested that objects and scenes may interact through shared visual feature information (Lowe et al., 2015), this dissertation will next explore object-scene interactions and test the hypothesis that interactions between an object and its background context will modulate the neural relationship of shared visual features by altering response patterns within regions of scene-selective cortex. Here, object and scene images will be combined from the previous experiment to form a new set of scenes which will examine whether the presence of an object with shared visual information (geometry and surface properties) influences neural representations within ventral visual cortex. By contrasting scene with object information, these experiments will also test the hypothesis that the processing of surface information in the PPA is domain specific to scenes, and thus the PPA would show greater activation when processing the surface properties of scenes compared with objects. Across the first two experiments, these explorations will additionally use multivoxel pattern analyses (MVPA), to examine if the processing of scene geometry and surface properties in the PPA are mediated by shared or distinct neuronal mechanisms. Finally, in Experiment 3, an fMR-adaptation approach will be used to obtain a sensitive measure of the relative weighting of geometry and surface properties in scene-perception. Consistent with previous predictions, this experiment will test the hypothesis that the PPA will show equal weighting (i.e., equivalent releases from adaptation) for the processing of the geometry and surface properties of a scene.

Second, how does feature diagnosticity and task context shape representations of these visual features, and higher-order descriptions of scenes? To address this question, this dissertation will present work which aims to investigate how geometry and surface properties contribute to neural representations of scenes across different scene categories by manipulating attention to these features separately (Chapter 3). In this chapter, neural activity in scene-
selective cortex will be examined while observers attend to either the layout or texture of natural and manufactured (otherwise referred to as manmade) scenes, either of which could change while the other was held constant. It is hypothesized that the PPA would show equal sensitivity to manipulations of both layout and texture in natural scenes, where textured zones and layout may be equally relevant for distinguishing scene identity. In manufactured scenes, however, it is hypothesized that the PPA would show less sensitivity to texture, relative to layout, as these scenes contain prominent horizontal and vertical structural components that can aid in the discrimination of scene identity. To isolate effects to the PPA, this research will also examine the modulation of brain activity in other areas of scene- and object-processing networks, and additionally localize a region of early visual cortex to examine if activation patterns observed in the PPA can be dissociated from activity in early visual areas. Building on previous research (Walther et al., 2009; 2011; Park et al., 2011; Kravitz et al., 2011), this work took advantage of both univariate and multivariate analyses to investigate previously unexplored questions of how task-dependent global scene features (i.e., attend to texture or layout) interact with high-level conceptual scene attributes (i.e., content: natural vs. manufactured scenes; and spatial boundary: open vs. closed scenes) to shape scene representation in human visual cortex.

Third, how do the processing of these visual features unfold across time to contribute to the perception of scene categories? To investigate how these properties interact and influence scene recognition within the temporal domain, this investigation applied a multifaceted approach using EEG to examine the time course of natural scene processing for six scene categories (beach; city; forest; highway; mountain; office) of colour photographs and line drawings during a simple memorization task (Chapter 4). This approach explored the temporal dynamics of scene perception across both basic-level (e.g., forest; city) and global-level (e.g., natural; manmade) categorizations using not only ERPs, but also a novel analysis which examined the correlations of within-category versus across-category discriminations at a precise temporal scale to investigate the time course of natural scene perception. Since previous research has shown that the distinction between global scene properties, such as the content of a scene (i.e., natural versus manmade), and the spatial boundary, or layout, of a scene (i.e., open versus closed) are important factors mediating scene perception and our ability to navigate through an environment (Oliva and Torralba, 2006; Park et al., 2011; Kravitz, Peng, and Baker, 2011), both global and basic-level categorizations of scene photographs and line drawings are examined to investigate how
visual feature information contributes to scene categorization. Building on previous research which has explored the spatial representations of scene photographs and line drawings (Walther et al., 2009; 2011), it is predicted that line drawings may be sufficient for scene categorization. By exploring the temporal dynamics of scene categorization, however, this examination further hypothesizes that the surface information provided from colour photographs may contribute to categorization in unique ways and diverge from line drawings, and this will be reflected through distinct neural markers which unfold across time. Specifically, if line drawings are sufficient for scene categorization, the structure preserved in line drawings may receive priority processing and therefore scene categorization may be resolved earlier than colour photographs.
Chapter 2

Neural Representation of Geometry and Surface Properties in Object and Scene Perception

In only the briefest of moments, the human visual system is able to draw on a broad array of cues to efficiently identify and navigate complex environments (Greene and Oliva, 2009). A common question has concerned how the human brain represents scene information to perform this feat. As discussed previously in the introduction, the spatial layout hypothesis, which emerged with the initial description of the PPA (Epstein and Kanwisher, 1998), posits that scene recognition can be understood through neural representations of the geometric structure, or spatial layout, of a scene, without the need to access surface information such as colour and texture. Yet surface property information is as abundant within the environment around us as geometry, and behavioural research has highlighted a role for colour and texture in mediating rapid scene recognition and understanding (Castelhano et al., 2008; Goffaux et al., 2005; Oliva and Schyns, 2000; Steeves et al., 2004; Renninger and Malik, 2004). The aim of Chapter 2 is thus to examine and compare the roles of geometry and surface properties in shaping underlying neural representations of the visual world, and therefore elucidate how these features contribute to scene recognition. Across 3 experiments I will focus on neural representations of these visual properties within the PPA, but additionally explore other regions of ventral visual cortex, including the scene-selective retrosplenial complex (RSC) and occipital place area (OPA), the object-selective lateral occipital (LO) area, and the face-sensitive fusiform face area (FFA) and occipital face area (OFA).

To explore the neural representation of geometry and surface properties, this chapter will examine these visual features in both object and scene perception. Since previous research has provided evidence that areas overlapping with the scene-selective PPA show sensitivity to

---

processing the surface properties of objects (Cant and Goodale, 2007; 2011), Experiment 1 will aim to replicate these findings on object perception and extend them to scene perception. Here, this investigation will test the hypothesis that the PPA will show equal weighting (i.e., equivalent levels of univariate activation) to the processing of the geometry and surface properties of a scene, but greater sensitivity to the surface properties of an object over its shape, when objects and scenes are presented separately. In Experiment 2, I will explore interactions between object and scene processing by examining neural representations of geometry and surface properties when objects and scenes are presented together, within the same context. In Experiment 3, an fMR-adaptation approach is used to obtain a sensitive measure of the relative weighting of geometry and surface properties solely within scene perception in PPA. In this experiment, it is predicted that equivalent releases from adaptation for variations in scene geometry or surface properties and an interaction (i.e., non-additivity) between the processing of these features will be observed. Non-additivity would imply that these processes involve similar underlying neuronal populations, which would further suggest that their representations are not independent.

Method

Observers

Thirty-six paid observers with normal or corrected-to-normal visual acuity were recruited from the University of Toronto community, consisting of ten paid observers (6 male; mean age 26.2 ± 4.92) in Experiment 1, twelve paid observers (6 male; mean age 25.83 years ± 3.61) in Experiment 2, and fourteen paid observers (6 male; mean age 24.21 ± 3.26) in Experiment 3. All Observers gave informed consent in accordance with the University of Toronto Ethics Review Board. One observer in Experiment 3 was removed prior to analyses due to excessive head motion (i.e., rotation and or translation in excess of 3 mm or 3°, respectively) which could not be motion-corrected within acceptable limits.
Stimuli and Procedure

E-Prime 2.0 (Psychology Software Tools, Pittsburgh, PA; Experiment 1; Experiment 2) and Matlab (MathWorks, Natick, MA; Experiment 3) were used to control stimulus presentation and collect behavioural responses. Images for all three experiments were rear-projected onto a screen in the MRI scanner (subtending 17.1° x 12.8 of visual angle), and observers viewed stimuli through a mirror mounted to the head coil directly above the eyes. In Experiment 1, stimuli were 512 unique full-colour 3-dimensional indoor scenes and objects rendered using Blender 2.0 software (Stichting Blender Foundation, Amsterdam; Fig. 1A) and created by varying a counterbalanced combination of scene-shape (circular; square), scene-texture (wood; brick), object-shape (circular; square), and object-texture (wood; brick). Textures were heterogeneous within a category (i.e., wood and brick), such that each category contained multiple exemplars of the same type of texture, as would be experienced in real-world environments (see Fig. 1A). One exemplar was used for each type of shape (i.e., circular and square; but see Experiment 3 for results when the number of texture and shape exemplars were matched). A blocked fMRI experimental paradigm was used wherein 24 images were presented in blocks of 20-s each. Each block was preceded by a 12-s fixation period and a 4-s written instruction to attend to changes in either the texture or shape of the scene or object in the ensuing block.

In each trial (12 per block, lasting 1666 ms each), two images were presented for 300 ms (separated by a 200-ms blank interval), and the task of the observers was to decide if the attended feature (shape or texture) of the stimulus (scene or object) was the same or different across the two images, responding during a 1166-ms period following the onset of the second image (via a response pad placed in the observer’s right hand). Each block contained an equal number of “same” and “different” trials. Observers were instructed to maintain central fixation and respond as quickly and as accurately as possible. Images in each condition were presented randomly within each block. Each observer took part in 5 runs (7 min 24 s each), and each run contained a unique and counterbalanced order of 12 stimulus blocks (i.e., three of each condition: scene-shape; scene-texture; object-shape; object-texture). Run order was randomized across observers and each condition (i.e., what was attended, shape or texture) was held constant per block. For Experiment 2, the procedure was identical, but the stimuli were combined to create new images...
containing both scenes and objects (Fig. 1B), with an object presented in the center of each scene.

In Experiment 3, 100 unique grayscale 3-dimensional indoor scenes were generated using Blender 2.0 software (Stichting Blender Foundation, Amsterdam; Fig 1C; Appendices Item 1). Stimuli were fully counterbalanced with 10 unique scene shapes, and 10 unique scene textures (i.e., variation in scene shape and scene texture were matched). Differences in low-level image features (luminance, contrast, and hue) were controlled using the SHINE toolbox (Willenbockel et al., 2010). A fast, event-related fMR-adaptation design was used, wherein 75 trials (6 s each) were presented in each run (7 min 42 s each), with a total of 5 runs per observer. Observers were asked to respond whether two consecutively displayed images were, as a whole, the “same” or different”, without attending to any particular stimulus dimension. Conditions examined included a “no-change” condition where neither feature changed (baseline), a “both-change” condition where both features changed, and two conditions where either feature could change independently of the other (“same-shape-different-texture”; “different-shape-same-texture”). In each trial, following an initial fixation of 1-s, two images were displayed (300 ms each) and were separated by an interstimulus interval of 800 ms. Following these images, observers responded during a 3600 ms response window (via a response pad placed in the observer’s right hand). Trial order was counterbalanced across five unique runs, and the order of these runs was counterbalanced across observers.

Localizer Scan

For Experiments 1 and 2, stimuli used to localize object-, scene-, and face-sensitive areas of cortex were photographs of various scenes, faces, common objects, and tile-scrambled images. Stimuli were presented in 16-s blocks of 32 images at a resolution of 375 x 375 pixels (7.8° x 7.8°), and were displayed for 400 ms each, with an interstimulus interval of 50 ms. Observers fixated on a centrally-presented black fixation cross, and were instructed to respond with a button press when the fixation cross changed from black to red (randomly occurring once or twice per stimulus block). There were 4 blocks for each stimulus category within a run, and there were two unique run orders. Sixteen-second long fixation periods were presented after each stimulus block. Each observer took part in three localizer runs (6-min 40-sec each). For
Experiment 3, stimuli used to localize object and scene-sensitive areas of cortex consisted of photographs of various scenes, faces, common objects, and phase-scrambled versions of the common objects. A single run consisted of presenting 4 blocks each of scenes, faces, intact objects, and phase-scrambled objects. Each stimulus block was 16-s long and contained 20 different images, each lasting 750 ms and followed by a 50 ms blank period. No images were repeated within or across blocks in a given run. To ensure attention to the displays, observers fixated at the center and detected a slight spatial jitter, occurring randomly in 1 out of every 10 images. Besides the stimulus blocks, there were also 8-s fixation blocks presented at the beginning, middle, and end of each run. Following Epstein and Kanwisher (1998), two unique and balanced run orders were used. Each run lasted 4 min and 40 s. All observers took part in 3 runs of this localizer.

**MRI Acquisition**

Scanning was performed at the Center for Addiction and Mental Health using a 3T GE Discovery MR750 whole-body MRI scanner equipped with an 8-channel head coil. T1-weighted anatomical images were acquired using a 3D SAG T1 BRAVO spiral pulse sequence [repetition time (TR), 6736 ms; echo time (TE), 3 ms; inversion time, 650 ms; flip angle 8°, 256 x 256 matrix size, 200 slices, 1 mm isovoxel]. For the functional runs, T2*-weighted images sensitive to blood oxygenation level-dependent (BOLD) contrasts were acquired using a spiral pulse sequence (64 x 64 matrix size; field of view 22 cm; TR 2000 ms; TE 30 ms; flip angle 60°; 200 volumes for the localizer runs in Experiments 1 and 2 and 140 volumes for the localizer runs in Experiment 3, 134 volumes for the main experimental runs in Experiments 1 and 2 and 231 volumes for the adaptation runs in Experiment 3). Thirty-one slices (3.4 mm x 3.4 mm x 5 mm, no gap) parallel to the anterior and posterior commissure line were collected in all functional runs.

**Univariate Data Analysis**

fMRI data were processed and analyzed using BrainVoyager QX 2.8 (Brain Innovation, Maastricht, the Netherlands). Data preprocessing included slice acquisition time correction, 3D
motion correction, temporal filtering (linear trend removal and high-pass filtering set at 3 cycles/run), and Talairach space transformation (Talairach and Tournoux, 1988). Data from the functional localizer was analyzed using a general linear model (GLM), accounting for hemodynamic response lag (Friston et al., 1994). Regions of interest (ROIs) can be seen in Figure 2. The PPA ROI was defined as a region in the collateral sulcus and parahippocampal gyrus (see Epstein and Kanwisher, 1998) whose activation was higher for scenes than for faces and objects (false discovery rate, q < 0.05; this threshold applies to all functional regions localized in individual observers; identified in all observers in all Experiments). In addition, the RSC (see Epstein and Higgins, 2005) and OPA (also known as transverse occipital sulcus; see Dilks et al., 2013) ROIs were functionally defined as regions in restrosplenial cortex–posterior cingulate–medial parietal cortex and transverse occipital cortex, respectively, whose activations were higher for scenes than for faces and objects (identified in eleven, and nine, observers, respectively, in Experiment 1, 7 and 9 observers in Experiment 2, and 12 and 13 observers in Experiment 3). In accordance with Grill-Spector et al. (2000), LO, a sub-division of the lateral occipital complex (LOC), was defined as a region in the lateral occipital cortex near the posterior inferotemporal sulcus, with activation higher for objects than for scrambled objects (identified in all observers in all experiments). The fusiform face area (FFA) was selected as a control region (as stimuli did not contain any faces) and following Kanwisher et al. (1997), this area was defined as a region in the extrastriate cortex whose activations were higher for faces than scenes or objects (identified in 11 observers in Experiment 1, all observers in Experiment 2, and 12 observers in Experiment 3). As an additional control region, the occipital face area (OFA) was defined as a region in the inferior occipital gyrus (Gauthier et al., 2000) whose activations were higher for faces than scenes or objects (identified in 9 observers in Experiment 1, 11 observers in Experiment 2, and 11 observers in Experiment 3).

Following the standard ROI-based analysis approach (Saxe et al., 2006), ROIs were overlaid onto the data from the main experiment and extracted time courses from each observer. Peak responses for each condition were obtained by collapsing the time courses for all of the conditions and then identifying the time point of greatest signal amplitude in the average response (Xu and Chun, 1997; Xu, 2010; Cant and Xu, 2012). This was done separately for each observer in each ROI, and the resultant peak responses were then averaged across all observers. The average levels of peak activation (measured in percent BOLD signal change from baseline
(fixation) for each condition across observers were subjected to a 2 (Experiment 1 and 2: Stimulus: object vs. scene; Experiment 3: Texture: same vs. different) x 2 (Experiment 1 and 2: Feature: texture vs. shape; Experiment 3: Shape: same vs different) repeated-measures ANOVA (SPSS, Chicago, IL, USA) for each ROI (PPA, RSC, OPA, LO, FFA, OFA). Planned pairwise comparisons (Bonferroni-corrected for multiple comparisons) were then conducted to compare the processing of texture and shape for both objects and scenes in each ROI in Experiments 1 and 2, and to compare same versus different conditions for scene texture and scene shape in Experiment 3. Left and right hemispheres were combined for each ROI in all analyses (see Hemisphere Analysis).

**Multivoxel Pattern Analysis (MVPA)**

*Support Vector Machine Classifiers*

Pattern classification was performed in Experiments 1 and 2 with a combination of in-house software (using Matlab) and the Princeton MVPA Toolbox for Matlab ([http://code.google.com/p/princeton-mvpa-toolbox/](http://code.google.com/p/princeton-mvpa-toolbox/)) using a Support Vector Machines (SVM) classifier (libSVM, [http://www.csie.ntu.edu.tw/~cjlin/libsvm/](http://www.csie.ntu.edu.tw/~cjlin/libsvm/)). The SVM model used a linear kernel function and a constant cost parameter, C=1, to compute a hyperplane that best separated the block/condition responses. To test the accuracy of the SVM classifiers, a “leave-one-run-out” N-fold cross-validation was used, in which a single fMRI run was reserved for classifier testing. This N-1 cross-validation procedure was performed until all runs were separately tested, and then averaged across N-iterations in order to produce a representative classification accuracy measure for each observer, ROI, and pattern discrimination (Duda et al., 1995).

**Multiclass and Pairwise Discriminations**

SVMs are designed for classifying differences between two stimuli and LibSVM (the SVM package implemented here) uses the so-called ‘one-against-one method’ for classification (Hsu and Lin, 2002). With the SVMs two complementary types of classification analyses were performed; one in which the multiple pairwise results were combined in order to produce
multiclass discriminations (distinguishing among all 4 condition types) and another in which the individual pairwise discriminations were examined and tested separately.

The multiclass discrimination approach allowed for an examination of the distribution of the classifier guesses through visualization of the resulting ‘confusion matrix’. In a confusion matrix, each row ($i$) represents the instances of the actual condition and each column ($j$) represents the predicted condition. Their intersection ($i, j$) represents the (normalized) number of times a given condition $i$ is predicted by the classifier to be condition $j$. Thus, the confusion matrix provides a direct visualization of the extent to which a decoding algorithm confuses (or correctly identifies) the different classes. All correct classifications are located in the diagonal of the matrix (with classification errors represented by non-zero values outside of the diagonal) and average decoding performance is defined as the mean across the diagonal. The values in each row sum to 1 (i.e., 100% classification). If decoding is at chance levels, then classification performance will be at $1/4 = 25\%$. For all multiclass discriminations, decoding significance was statistically assessed across observers (for each ROI and condition epoch) using one-tailed t-tests versus 25\% chance decoding. For pairwise discriminations, decoding significance was statistically assessed across observers using one-tailed t-tests versus 50\% chance decoding. Importantly, an FDR correction of $q \leq 0.05$ was applied to the pairwise comparisons based on the number of comparisons examined per ROIs, and for the multiclass discriminations based on the number of ROIs examined (Benjamini and Hochberg, 1995).

**Inputs to the SVM Classifier**

BOLD percent signal change values for each ROI provided inputs to the SVM classifier. The percent signal change response was computed from the time-course activity for the task-evoked responses with respect to the time-course of a run-based averaged baseline fixation value, for all voxels in the ROI. The baseline fixation window was defined as a time point prior to the 4-sec instruction period before each stimulus block (6 sec prior to block onset, averaged across all blocks within an experimental run). For the block-evoked activity, for each condition, the average of imaging volumes 3-10 (i.e., 6-20 sec) were extracted, which are time points encompassing the first peak of the hemodynamic response until the end of the experimental block. This windowed-average percent signal change classification approach corresponds with that used in recent work using the same technique (Gallivan et al., 2013; Gallivan et al., 2014;
Lowe et al., 2016). Following the extraction of each block’s activity, these values were rescaled between -1 and +1 for each voxel pattern within an ROI (Misaki et al., 2010).

**Behavioural data analysis**

Behavioural performance measures of accuracy and reaction time were recorded using E-Prime 2.0 software (Experiments 1 and 2) and Matlab (Experiment 3), and analyzed with SPSS, by performing a 2 (stimulus: object vs. scene) x 2 (feature: texture vs. layout) repeated-measures ANOVA for Experiments 1 and 2, and a 2 (texture: same vs. different) by 2 (shape: same vs. different) repeated-measures ANOVA for Experiment 3, with subsequent pairwise comparisons (all two-tailed and Bonferroni corrected) conducted based on a priori theoretical motivation (i.e., examining differences between shape and texture processing for objects and scenes separately in Experiments 1 and 2, and examining potential releases from adaptation when scene texture varied but scene shape was held constant, and when scene shape varied but scene texture was held constant, in Experiment 3). See Behavioural Analysis for results of the behavioural analyses for all three experiments.

**Experiment 1: Feature processing in objects and scenes separately**

In Experiment 1, observers viewed either indoor scenes or objects, separately (Figure 1A). Each scene and object could change along two dimensions: shape (circular; square) and texture (wood; brick). Given previous findings highlighting the influence of both surface properties (Cant and Goodale, 2007; 2011) and spatial layout (Epstein and Kanwisher, 1998) on activity in scene-selective cortex, I predict similar univariate neural activity between scene-texture and scene-shape conditions in the PPA (i.e., equal weighting). I further predict, based on previous results, increased univariate activity when attending to object-texture compared with object-shape in the PPA (Cant and Goodale, 2007; 2011), and vice versa in the object-selective LO, which has been shown to represent higher level object-shape information (Kourtzi and Kanwisher, 2001). To test the hypothesis that the PPA processes surface properties in a domain specific manner, I will examine neural representations of this visual feature across object and scene perception, predicting that the PPA will show increased activity to the surface properties of a scene over an object. Moreover, to test the prediction that geometry and surface properties are processed similarly within scene perception (i.e., are potentially mediated by shared neuronal
mechanisms), and distinctly within object perception, multivariate techniques are used to explore whether these features can be discriminated from one another in both scenes and objects. As the stimuli used here were not tailored to the known functional properties of FFA (Kanwisher et al., 1997) and OFA (Gauthier et al., 2000), these regions were used as controls (see Figure 2 for all ROIs).

Figure 1. Experimental Stimuli. (A) Examples of stimuli used in Experiment 1. Scenes and objects are defined by their shape (circular vs. square) and texture (wood vs. brick). Observers attended to the shape or texture of the object or scene, either of which could change while the other was held constant. (B) Examples of stimuli used in Experiment 2. The stimuli and procedure were identical to Experiment 1, with the exception that objects were placed within scenes. (C) Examples of the stimuli used in Experiment 3. Scenes could vary across 10 different shapes, and 10 different textures. In Experiment 3, observers attended only to overall changes across images, and did not attend directly to any one particular feature. For additional examples of the stimuli used in Experiment 3, see Appendix Item 1.

Figure 2. Regions of interest. Functionally defined ROIs are shown on the brains of two representative observers. Talairach coordinates for peak voxels of each ROI in observer one (PPA – FFA) and two (OFA) are as follows: LPPA, -29, -34, -10; RPPA, 27, -34, -5; LRSC, -15, -56, 14; RRSC, 14, -59, 17; LOPA, -41, -72, 14, ROPA, 32, -80, 19; LLO, -45, -80, 0; RLO, 39, -71, 5; LFFA, -43, -36, -17; RFFA, 41, -47, -13; LOFA, -36, -79, 4; ROFA, 43, -75, -3.
Results and Discussion

The analysis began by first examining the univariate response amplitudes (percent signal change compared to fixation) of ROIs for each condition (scene-shape; scene-texture; object-shape; object-texture) using 2 x 2 repeated-measures ANOVAs with stimulus (object; scene) and feature (texture; shape) as factors (Figure 3; collapsing left and right hemispheres for each region; this applies to all Experiments), and conducted simple main effects analyses to examine interactions between stimulus category and feature in each region. A significant main effect of stimulus for all regions except FFA was found (PPA: $F_{1,9} = 143.98, p < 0.001$; RSC: $F_{1,6} = 381.14, p < 0.001$; OPA: $F_{1,8} = 101.42, p < 0.001$; LO: $F_{1,9} = 22.08 p = 0.001$; FFA: $F_{1,9} = 0.36 p = 0.566$; OFA: $F_{1,9} = 16.17 p = 0.003$), and a main effect of feature for PPA ($F_{1,9} = 24.76, p = 0.001$), LO ($F_{1,9} = 11.00, p = 0.009$), and OFA ($F_{1,9} = 6.99 p = 0.027$), but not for the remaining regions (all $F$s < 4.59; all $p$s ≥ 0.061). Significant stimulus-by-feature interactions in all scene-selective regions were observed (PPA: $F_{1,9} = 35.03, p < 0.001$; RSC: $F_{1,6} = 9.12, p = 0.023$; OPA: $F_{1,8} = 20.63, p = 0.002$), but not in object-selective cortex (LO: $F_{1,9} = 0.06, p = 0.811$) or face-selective cortex (FFA: $F_{1,9} = 0.89 p = 0.369$; OFA: $F_{1,9} = 0.28 p = 0.609$).

These findings confirm a dissociation between object and scene processing across scene-selective and object-selective areas of cortex. In line with predictions, subsequent planned pairwise comparisons (two-tailed and Bonferroni-corrected; this applies to all Experiments) revealed no significant differences between scene-shape and scene-texture processing in the PPA ($t_9 = 0.17, p = 0.866$). Similarly, this was also observed in RSC ($t_6 = 1.75, p = 0.131$), yet higher BOLD response for scene-shape over scene-texture was observed in OPA ($t_8 = 3.09, p = 0.015$). Moreover, consistent with predictions, analysis of object-processing revealed significantly higher activation for object-texture over object-shape in the PPA ($t_9 = 12.61, p < 0.001$) but not RSC ($t_6 = 1.60, p = 0.170$), replicating previous findings (Cant and Goodale, 2007; 2011). These results further show higher activation for object-shape over object-texture in LO ($t_9 = 2.30, p = 0.047$), as expected. Finally, higher activation for object-texture over object-shape was also found in OPA ($t_8 = 3.15, p = 0.014$). This latter result was unexpected given the sensitivity of OPA to scenes, yet I speculate this finding may speak to the involvement of OPA in the processing of local elements, which may contain cues for scene recognition and navigation (Kamps et al., 2016a; 2016b).
Interactions across ROIs were investigated using a 2 (region) by 2 (stimulus) by 2 (feature) repeated-measures ANOVA. This analysis revealed significant region-by-stimulus interactions for PPA with RSC ($F_{1,6} = 45.43, p = 0.001$), LO ($F_{1,9} = 142.00, p < 0.001$), FFA ($F_{1,9} = 117.08, p < 0.001$), and OFA ($F_{1,9} = 134.46 p < 0.001$); for RSC with OPA ($F_{1,5} = 14.68, p = 0.012$), LO ($F_{1,9} = 98.15, p < 0.001$), FFA ($F_{1,6} = 160.90, p < 0.001$), and OFA ($F_{1,6} = 208.34 p < 0.001$); for OPA with LO ($F_{1,8} = 118.29, p < 0.001$), FFA ($F_{1,8} = 76.99, p < 0.001$) and OFA ($F_{1,8} = 122.78 p < 0.001$); for LO with FFA ($F_{1,8} = 19.93, p = 0.002$), and for FFA with OFA ($F_{1,9} = 27.81 p = 0.001$). Significant region-by-feature interactions were found for PPA with RSC ($F_{1,6} = 6.13, p = 0.048$), OPA ($F_{1,8} = 5.51, p = 0.047$), and LO ($F_{1,9} = 40.08, p < 0.001$); for RSC with OFA ($F_{1,6} = 6.11 p = 0.048$), for OPA with LO ($F_{1,8} = 17.77, p = 0.003$); and for LO with FFA ($F_{1,9} = 50.70, p < 0.001$) and OFA ($F_{1,9} = 21.67 p = 0.001$). A significant three-way interaction was found for PPA with LO ($F_{1,9} = 7.51, p = 0.023$), FFA ($F_{1,9} = 6.09, p = 0.036$), and OFA ($F_{1,9} = 12.34 p = 0.007$), and for OPA with OFA ($F_{1,8} = 5.52 p = 0.047$) These results are consistent with previous work demonstrating a functional dissociation between processing within scene-selective, object-selective, and face-selective cortex, but also within the scene-processing network itself (Cant and Goodale, 2007, 2011; Cant and Xu, 2012, 2015, 2016; Lowe et al., 2016).

Since a null result in a univariate analysis does not necessarily imply that a given region cannot distinguish between two properties, multivariate analyses were conducted to examine whether each of the ROIs could discriminate between stimulus category and feature. The aims of
these analyses were twofold. Firstly, to investigate the extent to which each ROI could successfully discriminate generally across all conditions (object shape, object texture, scene shape, scene texture), and secondly, to conduct a number of pairwise comparisons in order to directly explore whether these regions could successfully discriminate between the geometry and surface properties of objects and scenes. Multivoxel fMRI activity was first extracted and this analysis used linear SVM classifiers in each region to create confusion matrices representing the distribution of classifications (and misclassifications) across conditions (Figure 4A). These matrices demonstrate misclassifications in scene- and object-cortex are largely contained within the same stimulus category (e.g., scene-texture is more likely to be misclassified as scene-shape than object-texture). This analysis then examined the extent to which each condition could be decoded above chance (25%). These multiclass discriminations revealed classification accuracies that were significantly above chance for all ROIs (all $t$s $\geq$ 3.21; all $p$s $\leq$ 0.005; all $q$s $\leq$ 0.005), indicating that each region classifies conditions with above-chance accuracy (Figure 4B). To investigate these findings in greater detail, subsequent analyses were performed using planned pairwise-comparisons corrected for multiple comparisons using the FDR procedure ($q$; based on the number of comparisons per ROI) to examine whether shape and texture could be decoded at greater-than-chance accuracy (50%) when attending to either objects or scenes, separately (Figure 4C).
Findings revealed no significant decoding of scene-shape versus scene-texture in both PPA and RSC (both $t_s \leq 0.58$; both $p_s \geq 0.29$), replicating previous findings using real-world scene stimuli (Lowe et al., 2016), and thus provides additional evidence that scene-texture and scene-shape may be processed similarly in these regions. In contrast, significant decoding of scene-shape versus scene-texture was found in OPA ($t_s = 2.56, p = 0.017, q = 0.017$). Significant discrimination of these features was not observed in LO ($t_o = 1.16, p = 0.139$), nor in FFA ($t_o = 1.32, p = 0.110$), but it was found in OFA ($t_o = 3.44, p = 0.004, q = 0.005$). Continuing the analysis, significantly above-chance classification accuracy for object-shape versus object-texture was found in PPA ($t_o = 2.83, p = 0.010, q = 0.013$), providing mounting evidence consistent with previous univariate findings that the processing of these object features is dissociated in PPA (Cant and Goodale, 2007; 2011). Significant decoding of these object features was found in OPA ($t_s = 3.97, p = 0.002, q = 0.010$), but only marginally significant decoding of these features was found in LO ($t_o = 1.76, p = 0.056$), and no significant decoding was observed.
in RSC ($t_6 = 1.46, p = 0.098$). Significant decoding was found in both FFA ($t_9 = 2.21, p = 0.027, q = 0.045$) and OFA ($t_9 = 2.35, p = 0.022, q = 0.022$) across these conditions. Finally, to ensure a null finding in one of the main ROIs (e.g., PPA) could not be attributed to a problem with the classification procedure itself, control pairwise comparisons were conducted in which it was expected to find significantly above-chance classification. Specifically, this analysis examined the classification of scenes versus objects when holding stimulus feature constant, and found significant decoding of scene-shape versus object-shape for all regions (all $ts \geq 5.25$; all $ps \leq 0.001$; all $qs \leq 0.001$), as well as significant decoding of scene-texture versus object-texture for all regions (all $ts \geq 2.82$; all $ps \leq 0.01$; all $qs \leq 0.01$).

**Experiment 2: Feature processing in scenes containing an object**

Experiment 1 provides strong evidence that feature processing of a scene extends beyond spatial features to include surface properties such as texture, and that these features elicit similar neural activity in the PPA. In other words, the processing of scene texture and scene shape may be weighted equally in the PPA. Scenes rarely exist independently of objects, however, and previous research has indicated object-processing may have an interactive relationship with scene-processing (Joubert et al., 2007; Mullin et al., 2013) and that geometry and surface properties may influence this relationship (Lowe et al., 2015). To examine this interaction (i.e., the presence of a surrounding scene influencing object representation, and vice versa), Experiment 2 expanded upon the findings of Experiment 1 by examining the processing of shape and texture while observers viewed new images depicting an object placed within a scene (Figure 1B). This allowed an attempt to replicate the findings from Experiment 1, but importantly, to also test the prediction that shared visual object and scene features are not processed independently, but interact across stimulus categories. Specifically, since the surface properties of a scene can exhibit global precedence over the surface properties (but not the geometry) of an object within that scene (Lowe et al., 2015), results might show decreased sensitivity to object surface properties relative to the processing of object geometry in the PPA (e.g., equivalent activation for object texture and shape, or less activation for object texture). This would differ from the results in Experiment 1, in which objects were not presented within scenes.
Results and Discussion

Consistent with Experiment 1, univariate response amplitudes were examined in each region by conducting a 2 x 2 repeated-measures ANOVA with stimulus (object; scene) and feature (texture; shape) as factors (Figure 5), and found significant main effects of stimulus for all regions except a control region, FFA (PPA: $F_{1,11} = 88.40, p < 0.001$; RSC: $F_{1,10} = 16.74, p = 0.002$; OPA: $F_{1,10} = 110.83, p < 0.001$; LO: $F_{1,11} = 11.62, p = 0.006$; FFA: $F_{1,10} = 2.20, p = 0.169$; OFA: $F_{1,10} = 70.96, p < 0.001$), but only a main effect of feature for LO (PPA: $F_{1,11} = 2.25, p = 0.162$; RSC: $F_{1,10} = 0.004, p = 0.952$; OPA: $F_{1,10} = 0.334, p = 0.576$; LO: $F_{1,11} = 12.53, p = 0.005$; FFA: $F_{1,10} = 3.361, p = 0.097$; OFA: $F_{1,10} = 0.15, p = 0.711$). Interestingly, the stimulus-by-feature interaction was non-significant in all regions (all $Fs < 3.43$; all $ps ≥ 0.094$), which differs from the results of Experiment 1, where scenes and objects were presented separately. Subsequent pairwise comparisons found that activation levels for scene-shape and scene-texture did not differ in any scene selective region (PPA: $t_{11} = 0.35, p = 0.734$; RSC: $t_{10} = 0.19, p = 0.855$; OPA: $t_{10} = 1.00, p = 0.341$), and neither did activation for object-shape and object-texture in PPA ($t_{11} = 1.69, p = 0.118$). In contrast, levels of activity for these features were dissimilar in LO, with significantly higher activity observed when attending to object-shape compared with object-texture ($t_{11} = 2.53, p = 0.028$), as predicted. In line with predictions and consistent with the results of Experiment 1, these findings reveal that PPA exhibits similar BOLD responses across scene features. In contrast to the results of Experiment 1, however, these results showed equal sensitivity to the processing of object-texture and object-shape in PPA, which is consistent with the prediction regarding object-scene interactions.

![Figure 5. Univariate response amplitudes for Experiment 2. BOLD signal activation for all conditions (attend scene shape; attend scene texture; attend object shape; attend object texture) in each ROI. Data are represented as mean ± SEM.](image-url)
To explore these interactions more directly, a 2 (stimulus: object vs. scene) by 2 (feature: shape vs. texture) by 2 (Experiment: Experiment 1 vs. Experiment 2) mixed-design ANOVA was conducted for PPA across Experiments 1 and 2. Significant main effects of Experiment ($F_{1,20} = 5.19, p < 0.034$) and stimulus were observed ($F_{1,20} = 246.85, p < 0.001$), along with a significant stimulus-by-Experiment interaction ($F_{1,20} = 52.05, p < 0.001$). Moreover, a significant main effect of feature was found ($F_{1,20} = 16.41, p = 0.001$), but the feature-by-Experiment interaction was not significant ($F_{1,20} = 3.02, p = 0.098$). A significant main effect of feature was found ($F_{1,20} = 16.41, p = 0.001$), but the feature-by-Experiment interaction was not significant ($F_{1,20} = 3.02, p = 0.098$). A significant stimulus-by-feature interaction was found ($F_{1,9} = 18.12, p < 0.001$), and importantly, a significant three-way interaction was also observed ($F_{1,20} = 5.89, p = 0.025$), warranting further analyses. Subsequent pairwise comparisons revealed no significant differences between the representation of either the geometry ($t_{20} = 0.43, p = 0.671$) or surface properties ($t_{20} = 0.51, p = 0.611$) of a scene across experiments. In contrast, a significant difference between the representation of the geometry ($t_{20} = 4.98, p < 0.001$) and surface properties ($t_{20} = 4.13, p = 0.001$) of an object was found across experiments. Not surprisingly, higher response amplitudes in PPA for both of these object features was observed when the object was placed within a scene. But importantly, these results demonstrate that the representation in PPA changes when objects are placed within the context of a scene (specifically, the representation of object, but not scene, features changes).

Significant region-by-stimulus interactions were found for PPA with OPA ($F_{1,10} = 17.293, p = 0.002$), LO ($F_{1,11} = 67.07, p < 0.001$), FFA ($F_{1,10} = 126.45, p < 0.001$), and OFA ($F_{1,10} = 242.61 p < 0.001$); for RSC with OPA ($F_{1,9} = 21.25, p = 0.001$), LO ($F_{1,10} = 35.46, p < 0.001$), FFA ($F_{1,10} = 20.52, p = 0.001$), and OFA ($F_{1,9} = 58.17 p < 0.001$); for OPA with LO ($F_{1,10} = 61.81, p < 0.001$), FFA ($F_{1,9} = 74.95, p < 0.001$), and OFA ($F_{1,9} = 170.46 p < 0.001$), and for FFA with OFA ($F_{1,9} = 95.96 p < 0.001$). Significant region-by-feature interactions were found for PPA with LO ($F_{1,11} = 35.60, p < 0.001$), for RSC with LO ($F_{1,10} = 10.60, p = 0.009$), for OPA with LO ($F_{1,10} = 15.42, p = 0.003$), and for LO with FFA ($F_{1,10} = 25.89, p < 0.001$). A significant three-way interaction was found for RSC with OPA ($F_{1,9} = 12.24, p = 0.007$) and OPA with LO ($F_{1,10} = 7.69, p = 0.020$) and OFA ($F_{1,9} = 19.00 p = 0.002$). Taken together, the results from Experiments 1 and 2 demonstrate that scene-, object-, and face-sensitive regions of cortex process the same visual input in appreciably different ways, and speak to varying levels of functional specificity within ventral visual cortex (Cant and Goodale, 2007, 2011; Cant and Xu, 2012, 2015).
To further explore the processing of geometry and surface properties across ventral-visual cortex, a number of multivariate analyses were conducted consistent with Experiment 1. After creating confusion matrices (Figure 6A), multiclass discriminations revealed classification accuracies that were significantly above chance (25%) for all ROIs (all $ts \geq 4.70$; all $ps \leq 0.001$, all $qs \leq 0.001$), indicating that each region classifies conditions with above-chance accuracy (Figure 6B). To investigate these findings in greater detail, subsequent analyses were performed using planned pairwise-comparisons to examine whether shape and texture could be decoded at greater-than-chance accuracy (50%) when attending both objects and scenes, separately (Figure 6C). Critically, and consistent with the results of Experiment 1, no evidence was found for a significant discrimination of scene-shape and scene-texture conditions in both PPA and RSC (both $ts \leq 1.05$; both $ps \geq 0.159$), suggesting these features are processed similarly within these regions. In contrast, significant decoding of scene-shape versus scene-texture was found in OPA ($t_{10} = 2.09, p = 0.031, q = 0.031$), and similar results were observed in LO ($t_{11} = 2.15, p = 0.027, q = 0.027$). No difference across these conditions was found in a control region, FFA ($t_{10} = 0.88, p = 0.201$), but a difference was found in OFA ($t_{10} = 3.08, p = 0.006, q = 0.008$).

**Figure 6.** (Above) Confusion matrices for Experiment 2. (A) Confusion matrices (chance = 0.25) for Experiment 2 generated from multiclass discriminations showing the distribution of classification errors across all four conditions (SS = Scene Shape; ST = Scene Texture; OS = Object Shape; OT = Object Texture) for each ROI. The average classifier response proportions across observers are shown. When decoding is perfect, the confusion matrix will have a diagonal containing values of 1 and the rest of the matrix will be 0. Note that the average decoding performance (shown in Fig. 6c) is defined as the mean across the diagonal. To highlight differences in decoder performance, the matrices have been rescaled (rather than being scaled from 0-1). (Below) Multivariate results for Experiment 2. (B) Classification accuracy (chance = 25%; dashed line) of all four conditions for each ROI. (C) Decoding accuracy (chance = 50%; dashed line) for scene feature discriminations (shape vs. texture), object feature discriminations (shape vs. texture), shape category discriminations (scene vs. object) and texture category discriminations (scene vs. object) for each ROI. Data are represented as mean ± SEM. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, all $p$-values shown have been FDR-corrected.
We next examined the decoding of object-shape versus object-texture, and found significantly above-chance classification accuracy in PPA ($t_{11} = 4.50, p < 0.001, q = 0.001$), despite similar univariate response amplitudes. These results confirm a dissociation of these object features in PPA consistent with previous findings (Cant and Goodale, 2007; 2011), and further highlight the advantage of an approach using both univariate and multivariate analyses: A null result in univariate response does not necessarily imply a null multivariate result. Thus, examining both enabled us to better characterize the relationship between the weighting of visual features and the degree to which they are represented by shared or distinct neuronal populations. Significant decoding of these object features was found in OPA ($t_{10} = 3.68, p = 0.002, q = 0.003$) and LO ($t_{11} = 4.10, p < 0.001, q = 0.001$), but not in RSC ($t_{10} = 1.60, p = 0.070$), FFA ($t_{10} = 0.64, p = 0.267$), and OFA ($t_{10} = 0.84, p = 0.210$). Finally, significant decoding of scene-shape versus object-shape was found in all regions (all $t$s $\geq 3.90$; all $p$s $\leq 0.002$, all $q$s $\leq 0.003$) except FFA ($t_{11} = 1.86, p = 0.046, q = 0.093$). For the decoding of scene-texture versus object-texture, significant classification in all regions was observed (all $t$s $\geq 3.72$; all $p$s $\leq 0.002$, all $q$s $\leq 0.004$). These latter sets of results demonstrate that any null result in decoding accuracy (e.g., scene shape vs. scene texture in PPA) cannot be explained by imperfections in the classification algorithm itself.

**Experiment 3: fMR-adaptation of scene features**

The results of the first two experiments suggested that the processing of geometry and surface properties are weighted equally in scene representation, and are possibility mediated by shared neuronal mechanisms. To provide an additional test of this hypothesis, this experiment utilized a fast event-related fMR-adaptation paradigm, which allowed examination of the neural representation of one feature (e.g., scene texture), independent of changes in a second feature (e.g., scene shape). If scene shape and scene texture are not weighted equally in the PPA, then changing each feature in isolation should result in significantly different releases from adaptation (compared to a no feature change baseline). If these features are weighted equally, however, then I expect to observe equivalent releases from adaptation. Based on the results from the previous experiments, the latter scenario is predicted. Moreover, by examining potential interactions between scene shape and scene texture processing, this investigation is able to assess whether these features are represented additively or non-additively. Given previous results suggesting
equal weighting of these features in the PPA (mediated by potentially shared underlying neuronal mechanisms), an interaction (i.e., non-additivity) between the processing of these features is expected, which would imply that their representations are not independent in the PPA.

Furthermore, previous work has shown that goal-states and attentional task demands directly influence activity in ventral visual cortex (Harel et al., 2014). In addition to this potential influence, there is strong evidence to suggest that scene perception may be impacted by lower-level stimulus properties such as colour (Oliva and Schyns, 2000; Steeves et al., 2004; Goffaux et al., 2005; Castelhano and Henderson, 2008). Together, these findings raise the possibility that the results of the previous experiments, which utilized full-colour stimuli and manipulated attention to particular stimulus features, may be partially explained by these factors. Thus, this experiment controlled for these potential caveats by examining the response properties of ventral-stream regions when observers were not explicitly attending to a particular stimulus dimension, but instead were performing a more general same-different judgement (Figure 1C). Experiment 3 focuses this investigation exclusively on scenes to avoid potential modulation by interactions between objects and scenes (see Experiment 2). Additionally, this experiment controlled for low-level image properties by using grayscale scenes and then processing these images by using the SHINE toolbox (Willenbockel et al., 2010), which equates low-level image attributes across images by normalizing luminance, contrast, and hue.

Finally, it is possible that the equal univariate activation for scene shape and scene texture in the PPA in Experiments 1 and 2 resulted from unmatched variation across these features. That is, because only two instances of scene shape were used (i.e., round vs. square), but many instances of scene texture (i.e., many different types of brick and wood textures), blocks where observers attended to scene shape would result in more adaptation (i.e., less activation) compared with blocks where they attended to scene texture. If the representation of these features is not equally weighted in the PPA (i.e., attending to scene shape normally elicits greater activation then attending to scene texture), then over adapting scene shape compared with scene texture would give the appearance of equivalent univariate activation and thus equal weighting. To investigate this possibility, variations across scene geometry and texture were matched by using ten different scene-shapes that could each be rendered in ten different scene-
textures. If equivalent releases from adaptation are observed when changing scene shape and scene texture independently, then the results of Experiments 1 and 2 cannot be explained by unmatched variation across these features.

Results and Discussion

fMR-adaptation to geometry and surface properties was analyzed using a 2 x 2 repeated-measures ANOVA with shape (same; different) and texture (same; different) as factors (Figure 7). Importantly, this design allowed us to examine the interaction between scene shape and texture processing, but similar results (with regard to releases from adaptation for variations in scene shape or scene texture) were obtained when a one-way repeated measures ANOVA was used with the factor condition (no change vs. same shape, different texture vs. different shape, same texture vs. both change) to analyze the data. Results from the 2 x 2 analysis revealed a significant main effect of shape for PPA \((F_{1,12} = 8.22, p = 0.014)\), RSC \((F_{1,11} = 8.13, p = 0.016)\), and LO \((F_{1,12} = 16.59, p = 0.002)\), but not for OPA \((F_{1,12} = 1.04, p = 0.328)\) or face-selective regions (FFA: \(F_{1,11} = 1.28, p = 0.283\); OFA: \(F_{1,10} = 0.01 p = 0.928\)). In addition to a main effect of shape, a main effect of texture was found in PPA \((F_{1,12} = 8.21, p = 0.014)\), but not the remaining regions (all \(Fs < 3.90; all ps \geq 0.074\)). Significant shape-by-texture interactions were found in PPA \((F_{1,12} = 14.55, p = 0.002)\), OPA \((F_{1,12} = 16.90, p = 0.001)\), and FFA \((F_{1,11} = 8.00, p = 0.016)\), but not the remaining regions (all \(Fs < 4.56; all ps \geq 0.058\)). Next, planned pairwise comparisons were conducted to examine fMR-adaptation effects across conditions. Compared to the no-change condition, the same-shape-different-texture condition resulted in a release from adaptation for all regions (all \(ts \geq 2.21; all ps \leq 0.047\)), except OFA \((t_{10} = 1.79, p = 0.104)\). Similarly, the different-shape-same-texture condition showed a release when compared to the no-change condition in all regions except FFA and OFA (PPA: \(t_{12} = 5.61, p < 0.001\); RSC: \(t_{11} = 2.69, p = 0.021\); OPA: \(t_{12} = 3.40, p = 0.005\); LO: \(t_{12} = 4.20, p = 0.001\); FFA: \(t_{11} = 1.46, p = 0.173\), OFA: \(t_{10} = 1.63, p = 0.135\).
Interactions between regions were analysed using a 2 (region) by 2 (shape) by 2 (texture) repeated-measures ANOVA. Significant region-by-shape interactions were found for PPA with FFA ($F_{1,11} = 7.16, p = 0.022$), for RSC with FFA ($F_{1,11} = 9.15, p = 0.012$), for OPA with FFA ($F_{1,11} = 7.01, p = 0.022$) and for LO with FFA ($F_{1,11} = 7.36, p = 0.020$). Significant region-by-texture interactions were found for PPA with RSC ($F_{1,11} = 4.91, p = 0.049$) and LO ($F_{1,12} = 6.80, p = 0.023$), and for RSC with FFA ($F_{1,11} = 4.90, p = 0.049$). No significant three-way interactions were found.

It should be noted that despite controlling for multiple low-level visual properties in the stimulus set (e.g., colour, luminance, contrast), it is likely that some low-level differences persist. For example, changing scene shape in the stimuli, while keeping scene texture constant, produced low-level changes in monocular depth cues from surface properties. Thus, neural responses in the different-shape-same-texture condition likely result from a combination of differences in scene shape and differences in low-level depth cues from texture. The results of this experiment demonstrated equivalent releases from adaptation across all ‘change’ conditions, however, and the adaptation results are consistent with the results from the previous two experiments. Together, this makes it unlikely that differences in low-level properties alone would account for a significant proportion of the neural response in the different-shape-same-texture condition, especially since numerous studies have found robust sensitivity of the PPA to variations in scene shape (e.g., Epstein and Kanwisher, 1998). Nevertheless, it is possible that changes to global scene shape and layout in previous studies were correlated with changes to

Figure 7. fMR-Adaptation response amplitudes for Experiment 3. Levels of adaptation across conditions (no change; same-shape-different-texture; different-shape-same-texture; both change) for each ROI. Data are represented as mean ± SEM.
low-level monocular depth cues from surface properties, as they were in the different-shape-
same-texture condition here. Future studies are thus required to better understand the perceptual
and neural impact of covariations in different scene features. I would contend, however, that the
contribution of depth-from-texture information to these results is not inconsistent with the
general conclusion that both geometry and surface properties contribute to scene representation
in the PPA.

**Hemisphere Analysis**

An examination of differences in response amplitudes across hemispheres for each of the
ROIs used a 2 (hemisphere: left vs. right) by 2 (stimulus: object vs. scene) by 2 (feature: shape
vs. texture) repeated-measures ANOVA for Experiment 1 and Experiment 2. In Experiment 1, no
main effect of hemisphere, nor an interaction between hemisphere and either stimulus or feature
was found in PPA (all \( F_s \leq 2.05, \) all \( p_s \geq 0.186 \)), RSC (all \( F_s \leq 5.20, \) all \( p_s \geq 0.063 \)), LO (all \( F_s \leq
1.58, \) all \( p_s \geq 0.240 \)), and FFA (all \( F_s \leq 2.92, \) all \( p_s \geq 0.122 \)). In OPA, no significant main effect
of hemisphere was found \((F_{1,8} = 4.65, p = 0.385)\), nor a hemisphere-by-stimulus interaction \((F_{1,8}
= 0.29, p = 0.608)\), but a significant hemisphere-by-feature interaction was observed \((F_{1,8} = 9.38,
p = 0.016)\). Similarly, in Experiment 2, no main effect of hemisphere, nor an interaction was
found in PPA (all \( F_s \leq 3.94, \) all \( p_s \geq 0.073 \)), RSC (all \( F_s \leq 2.67, \) all \( p_s \geq 0.134 \)), LO (all \( F_s \leq
4.86, \) all \( p_s \geq 0.05 \)), and FFA (all \( F_s \leq 0.12, \) all \( p_s \geq 0.732 \)). In OPA, no significant main effect of
hemisphere was found \((F_{1,10} = 0.83, p = 0.385)\), nor a hemisphere-by-stimulus interaction \((F_{1,10} <
0.0, p = 0.954)\), but again a significant hemisphere-by-feature interaction was observed \((F_{1,10} =
7.28, p = 0.022)\).

In Experiment 3, a 2 (hemisphere: left vs. right) by 2 (shape: same vs. different) by 2
(texture: same vs. different) repeated-measures ANOVA was used to examine hemispheric
differences. No main effect of hemisphere, nor an interaction between hemisphere and either
shape or texture was found in PPA (all \( F_s \leq 2.37, \) all \( p_s \geq 0.150 \)), RSC (all \( F_s \leq 5.16, \) all \( p_s \geq
0.305 \)), OPA (all \( F_s \leq 2.15, \) all \( p_s \geq 0.168 \)), and FFA (all \( F_s \leq 2.71, \) all \( p_s \geq 0.128 \)). In LO, no
main effect of hemisphere was observed \((F_{1,12} = 0.31, p = 0.590)\), but a hemisphere-by-shape
interaction \((F_{1,12} = 5.93, p = 0.031)\), and a hemisphere-by-texture interaction \((F_{1,12} = 5.02, p
= 0.045)\), was found. To examine these two interactions in greater detail, a 2 (shape: same vs.
different) by 2 (texture: same vs. different) repeated-measures ANOVA was performed separately in each hemisphere of LO. In right LO, a significant main effect of shape \((F_{1,12} = 8.31, p = 0.014)\) was found, but the main effect of texture \((F_{1,12} = 3.94, p = 0.071)\) and the shape-by-texture interaction were not significant \((F_{1,12} = 1.70, p = 0.216)\). In left LO, the same pattern of results was observed, showing a significant main effect of shape \((F_{1,12} = 13.14, p = 0.003)\), but not a significant main effect of texture \((F_{1,12} = 0.047, p = 0.833)\), and no significant interaction between the two \((F_{1,12} = 1.92, p = 0.191)\). Thus, hemispheres were collapsed across each region to examine the main hypotheses.

**Behavioural results**

Behavioural accuracy and reaction time in Experiments 1 and 2 were analyzed using a 2 x 2 repeated-measures ANOVA with main effects of stimulus (object; scene) and feature (shape; texture; **Table 1**). For Experiment 1, when examining accuracy, significant main effects of both stimulus \((F_{1,9} = 30.54, p < 0.001)\) and feature \((F_{1,9} = 17.26, p = 0.002)\), and a significant interaction \((F_{1,9} = 14.37, p = 0.004)\), were present. Similarly, for reaction time, a significant main effect of stimulus \((F_{1,9} = 87.62, p < 0.001)\) and feature \((F_{1,9} = 62.24, p < 0.001)\), and a significant interaction \((F_{1,9} = 28.96, p < 0.001)\), was found. Post-hoc pairwise comparisons (two-tailed and Bonferroni-corrected) found significant differences between scene-shape and scene-texture for both accuracy \((t_9 = 4.47, p = 0.009)\) and reaction time \((t_9 = 5.87, p = 0.001)\), and object-shape and object-texture for only reaction time \((t_9 = 7.41, p < 0.001)\). Behavioural accuracy and reaction time were next examined across conditions for Experiment 2. In line with the results of Experiment 1, when examining accuracy, a significant main effect of both stimulus \((F_{1,11} = 34.15, p < 0.001)\) and feature \((F_{1,11} = 112.483, p < 0.001)\), and a significant interaction \((F_{1,11} = 25.99, p < 0.001)\), was found. For reaction time, significant main effects of stimulus \((F_{1,11} = 47.67, p < 0.001)\), feature \((F_{1,11} = 30.58, p < 0.001)\), and a significant interaction \((F_{1,11} = 7.49, p = 0.019)\), were observed. Post-hoc pairwise comparisons (two-tailed and Bonferroni-corrected) found significant differences between scene-shape and scene-texture for both accuracy \((t_{11} = 11.36, p < 0.001)\) and reaction time \((t_{11} = 3.30, p = 0.007)\), and between object-shape and object-texture for both accuracy \((t_{11} = 3.13, p = 0.010)\) and reaction time \((t_{11} = 6.02, p < 0.001)\).
Table 1
Accuracy (percent correct) and reaction time (ms) for each condition in each Experiment (1-3)

<table>
<thead>
<tr>
<th>Experiment 1</th>
<th>Accuracy</th>
<th>Reaction Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Object Texture</td>
<td>96.14 ± 1.05</td>
<td>562.30 ± 18.99</td>
</tr>
<tr>
<td>Object Shape</td>
<td>97.41 ± 0.84</td>
<td>511.92 ± 15.98</td>
</tr>
<tr>
<td>Scene Texture</td>
<td>88.79 ± 1.07</td>
<td>581.54 ± 17.49</td>
</tr>
<tr>
<td>Scene Shape</td>
<td>96.22 ± 1.17</td>
<td>563.47 ± 18.24</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Experiment 2</th>
<th>Accuracy</th>
<th>Reaction Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Object Texture</td>
<td>94.99 ± 1.01</td>
<td>573.94 ± 20.70</td>
</tr>
<tr>
<td>Object Shape</td>
<td>97.92 ± 0.64</td>
<td>526.22 ± 22.40</td>
</tr>
<tr>
<td>Scene Texture</td>
<td>85.16 ± 1.38</td>
<td>594.13 ± 22.78</td>
</tr>
<tr>
<td>Scene Shape</td>
<td>94.94 ± 1.08</td>
<td>568.96 ± 20.11</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Experiment 3</th>
<th>Accuracy</th>
<th>Reaction Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Change</td>
<td>93.35 ± 1.49</td>
<td>758.93 ± 39.18</td>
</tr>
<tr>
<td>Same Shape Different Texture</td>
<td>92.22 ± 1.33</td>
<td>747.86 ± 35.39</td>
</tr>
<tr>
<td>Different Shape Same Texture</td>
<td>94.44 ± 1.20</td>
<td>766.93 ± 33.11</td>
</tr>
<tr>
<td>Both Change</td>
<td>95.14 ± 1.63</td>
<td>736.06 ± 33.13</td>
</tr>
</tbody>
</table>

All values represent mean ± SE

While these results indicate differences in task difficulty across conditions, previous research exploring the impact of task performance on neuronal representations has found clear evidence for a dissociation between manipulations of task difficulty and processing in ventral visual areas, including the PPA (Xu et al., 2007). This suggests that these regions may represent stimulus-specific processing uncontaminated by differences in task difficulty. Consistent with this finding, patterns of behavioural responses in the present study were not reflected in univariate response amplitudes in both Experiments 1 and 2. For example, object shape and object texture differed in accuracy (object shape > object texture: Experiment 2) and reaction time (object shape < object texture: Experiments 1 and 2), yet these two conditions demonstrated opposite patterns of activation across the PPA (object shape < object texture) and LO (object shape > object texture). It is also not likely that the results are accounted for by different cognitive strategies across the conditions, since there is no evidence for a speed/accuracy trade-off in the behavioural results (i.e., in both Experiments 1 and 2, the fastest condition is also the
most accurate). Moreover, scene shape and scene texture differed in both behavioural measures in Experiments 1 and 2, yet they did not differ in neural response amplitude in the PPA in both experiments. Thus, it is difficult to conceive how differences in behavioural measures alone contributed to the neural results observed in Experiments 1 and 2.

For Experiment 3, an examination of behavioural accuracy and reaction time across conditions using a 2 x 2 repeated-measures ANOVA with shape (same; different) and texture (same; different) as factors found a significant main effect for shape ($F_{1,12} = 14.25$, $p = 0.003$), but not texture ($F_{1,12} = 0.072$, $p = 0.793$), and a non-significant interaction ($F_{1,12} = 0.872$, $p = 0.369$), when examining accuracy. For reaction time, no significant main effects nor an interaction were found (all $F_s < 2.76$, $p > 0.122$). As per previous experiments, planned pairwise comparisons (two-tailed and Bonferroni-corrected) were subsequently conducted and these comparisons revealed no significant differences in accuracy for both the no-change compared with the same-shape-different-texture condition ($t_{12} = 0.772$, $p = 0.455$) and the no-change compared with the different-shape-same-texture condition ($t_{12} = 1.00$, $p = 0.339$). Similar results were found when examining reaction time (no-change vs. same-shape-different-texture condition: $t_{12} = 0.52$, $p = 0.611$; no-change vs. different-shape-same-texture condition: $t_{12} = 0.38$, $p = 0.712$). Taken together, the behavioural and neural results of Experiments 1, 2, and 3 are in line with previous research indicating that activity in the PPA is impervious to manipulations of task difficulty as neural results are consistent across three experiments, despite evidence of differences in task difficulty in Experiments 1 and 2, but not 3. However, it is not prudent to completely dismiss potential interactions between neural and behavioural results in the present study, and subsequent studies should investigate these questions.

**Summary**

The findings presented in Chapter 2 demonstrate that the PPA responds just as strongly to changes in the surface properties of a scene as it does to changes in spatial structure. Neural responses to these scene features could not be discriminated from one another in the PPA, despite significant discrimination of these features in object perception, suggesting these features may share similar underlying neuronal mechanisms. Greater responses to scene-texture compared with object-texture in the PPA were also observed, regardless of whether scenes and
objects were presented separately or together, suggesting that the processing of surface properties within this region may be mediated by domain-specific, rather than domain-general mechanisms. Interestingly, while the PPA showed greater sensitivity to processing texture compared with shape in object perception when objects were viewed independently of scenes, an interaction across object and scene perception altered this relationship, wherein equal sensitivity to these properties was found when objects were perceived in the context of a scene. To conclude, these results suggest that representations of scene geometry and surface properties are weighted equally in the PPA, with their processing potentially mediated by similar underlying neuronal mechanisms.
Chapter 3
Feature Diagnosticity and Task Context Shape Activity in Human Scene-selective Cortex

The results of Chapter 2 have provided evidence that not only is surface information represented within scene-selective areas of visual cortex, but that this information may be just as important as geometry in shaping neural representations of the visual environment. These results challenge the view that neural mechanisms underlying scene recognition are driven by spatial features alone, and support a more flexible view of how scenes are represented in the brain. To better understand these representations, however, it is not only necessary to understand the contributions of individual features, but also how these features converge and align with behavioural goals to contribute to the formation of scene identity. Disentangling this information has been a challenge, as these features rarely exist independently, and may inform scene identity through complex interactions which vary according to scene category and task demands.

Early research exploring diagnostic visual features in the recognition of objects revealed a primary role for edge-based information (i.e., structure), suggesting surface characteristics such as colour and texture play only a secondary role in object recognition (Biederman and Ju, 1988). Research has since extended support for edge-based determinants of visual object recognition to scene perception (Delorme et al., 2000; Walther et al., 2011; Walther and Shen, 2014), yet a growing body of work suggests diagnostic surface characteristics such as colour and texture are instrumental in mediating early-stage scene gist processing that is responsible for successful scene recognition (Schyns and Oliva, 1994; Oliva and Schyns, 2000; Goffaux et al., 2005; Steeves et al., 2004; Castelhano and Henderson, 2008). Given these differences, the interplay between surface properties and structural features as determinants for scene recognition is currently unclear. One framework for scene perception that was discussed previously in the

---

Introduction, and which may reconcile these differences, proposes that the recognition of complex visual scenes can be understood through interactions between perceptually available information and categorization demands (Oliva and Schyns, 1997). This recognition framework involves a model of feature diagnosticity, or the idea that certain visual cues are utilized for certain types of categorizations, which may interact with behavioural goals and available visual information for the purposes of recognition. In this framework, diagnostic visual features emerge as a function of their usefulness in defining the identity of a scene, and the task demands placed on the observer. Since scenes are highly complex and can vary greatly from one scene to the next, diagnosticity of visual information may be flexible and dynamic. The present chapter thus aims to investigate the influence of diagnostic surface- and edge-based visual features on neural scene processing across a range of scene categories.

In light of the importance of geometry and surface information in scene perception (Chapter 2), and potential differences in the relevance of these features for categorizing different scenes, Chapter 3 examined neural activity in scene-selective cortex while observers attended to either the layout or texture of natural and manufactured scenes, either of which could change while the other was held constant. In this study, multiple categories of scenes were examined, including scene categories describing global spatial boundary information (open versus closed) and scene content (natural versus manufactured), as well as different basic-level categories of scenes (cities, caves, indoor rooms, deserts). Here, it was hypothesized that the PPA would show equal sensitivity to manipulations of both layout and texture in natural scenes, where textured zones and contour information from layout both may be important for distinguishing scene identity (Oliva and Torralba, 2001). In manufactured scenes, however, it was hypothesized that the PPA would show less sensitivity to texture, relative to layout, as these scenes contain prominent horizontal and vertical structural components that can aid in the discrimination of scene identity. To isolate effects to the PPA and compare and contrast these effects to other regions of ventral visual cortex, this investigation also examined the modulation of brain activity in other areas of scene-, object-, and face-processing networks (RSC, OPA, LO, FFA), and additionally localized a region of early visual cortex (EVC) to examine if activation patterns observed in the PPA can be dissociated from activity in early visual areas.
Method

Observers

Twelve paid observers (6 male; mean age 27.4 ± 3.8 years) with normal or corrected-to-normal visual acuity were recruited from the University of Toronto community. Observers gave informed consent in accordance with the University of Toronto Ethics Review Board.

Stimuli and procedure

Stimuli were grayscale photographs from four different scene categories devoid of foreground objects to avoid interference (see Davenport and Potter, 2004; Joubert et al., 2007), and created by varying features of spatial boundary (open vs. closed) and scene content (natural vs. manufactured; Figure 8) (Oliva and Torralba, 2001). After selection of four scene categories, twelve unique structural arrangements (i.e., layouts) were selected for each category, and twelve appropriate textures were applied to the dominant surface of each layout (mapped onto scene gradient and depth using Adobe Photoshop CS3), yielding 144 unique images per scene category (12 layouts/category x 12 textures/layout x 4 scene categories = 576 total images). E-Prime 2.0 (Psychology Software Tools, Pittsburgh, PA) was used to control stimulus presentation and collect behavioural responses. Images were rear-projected onto a screen in the MRI scanner at a resolution of 500 x 500 pixels (subtending 10.4° x 10.4° of visual angle), and observers viewed stimuli through a mirror mounted to the head coil directly above the eyes. A blocked fMRI experimental paradigm was used, wherein sixteen images from a single scene category were presented in blocks of 16-s each. Each block was preceded by a 12-s fixation period and a 4-s written instruction to attend to changes in either the texture or layout of the scenes in the ensuing block.
In each trial (8 per block, lasting 2 sec each), two scenes were presented for 300 ms (separated by a 200 ms blank interval), and the task of the observers was to decide if the attended feature (i.e., layout or texture) was the same or different across the two images, responding during a 1.5-s period following the onset of the second image (via a response pad placed in the observer’s right hand). Each block contained an equal number of ‘same’ and ‘different’ trials. Observers were instructed to maintain central fixation and respond as accurately as possible, placing no emphasis on fast response times to help encourage accurate performance. Images from a single scene category were presented randomly within each block, and each image could be repeated only once per observer. Each observer took part in 8 runs (4-min 28-sec each). Each
run contained a unique and counterbalanced order of 8 different stimulus blocks (i.e., 8 different conditions: attend to texture or layout in each of the four scene categories). Run order was randomized across observers, and scene category was held constant per block.

_Localizer scan_

Stimuli used to localize object-, scene-, and face-sensitive areas of cortex, as well as early visual cortex, were photographs of various scenes, faces, common objects, and tile-scrambled images. Stimuli were presented in 16-s blocks of 32 images at a resolution of 375 x 375 pixels (7.8° x 7.8°), and were displayed for 400 ms each, with an interstimulus interval of 50 ms. Observers fixated on a centrally-presented black fixation cross, and were instructed to respond with a button press when the fixation cross changed from black to red (randomly occurring once or twice per stimulus block). There were 4 blocks for each stimulus category within a run, and there were two unique run orders. Each observer took part in three localizer runs (6-min 40-sec each).

_MRI acquisition_

Scanning was performed at the Center for Addiction and Mental Health using a 3T GE Discovery MR750 whole-body MRI scanner equipped with an 8-channel head coil. T1-weighted anatomical images were acquired using a 3D SAG T1 BRAVO spiral pulse sequence [repetition time (TR), 6736 ms; echo time (TE), 3 ms; inversion time, 650 ms; flip angle 8°, 256 x 256 matrix size, 200 slices, 1 mm isovoxel]. For the functional runs, T2*-weighted images sensitive to blood oxygenation level-dependent (BOLD) contrasts were acquired using a spiral pulse sequence (64 x 64 matrix size; field of view 22 cm; TR 2000 ms; TE 30 ms; flip angle 60°; 200 volumes for the localizer runs, 134 volumes for the main experimental runs). Thirty-one slices (3.4 mm x 3.4 mm x 5 mm, no gap) parallel to the anterior and posterior commissure line were collected in all functional runs.
Univariate data analysis

fMRI data were processed and analyzed using BrainVoyager QX 2.8 (Brain Innovation, Maastricht, the Netherlands; see Chapter 2 for details). Regions of interest (ROIs) can be seen in Figure 9. In accordance with Epstein and Kanwisher (1998), the PPA ROI was defined as a region in the collateral sulcus and parahippocampal gyrus whose activation was higher for scenes than for faces and objects (false discovery rate, q < 0.05; this threshold applies to all functional regions localized in individual observers; identified in all twelve observers). In addition, in accordance with Epstein and Higgins (2007) and Dilks et al. (2013), retrosplenial complex (RSC) and the occipital place area (OPA; also known as transverse occipital sulcus) ROIs were functionally defined as regions in retrosplenial cortex–posterior cingulate–medial parietal cortex and transverse occipital cortex, respectively, whose activations were higher for scenes than for faces and objects (identified in eleven, and nine, observers, respectively). In accordance with Grill-Spector et al. (2000), the lateral occipital area (LO) was defined as a region in the lateral occipital cortex near the posterior inferotemporal sulcus, with activation higher for objects than for tile-scrambled objects (identified in all twelve observers). Early visual cortex (EVC) was defined as a retinotopic region around the calcarine sulcus with activation higher for scrambled objects than intact objects (MacEvoy and Epstein, 2011; Cant and Xu, 2016; identified in eleven observers).
Following the standard ROI-based analysis approach (see Chapter 2 for details), the average levels of peak activation (measured in percent BOLD signal change from baseline fixation) for each condition across observers were subjected to a 6 (ROI: PPA, RSC, OPA, LO, FFA, EVC) x 2 (spatial boundary: open vs. closed) x 2 (content: natural vs. manufactured) x 2 (task: texture vs. layout) repeated-measures ANOVA (SPSS, Chicago, IL, USA). This analysis revealed differences in activation across ROIs, so further analyses were conducted on each ROI separately. Moreover, subsequent analyses revealed no significant differences in activation between open and closed scenes in scene-selective cortex, so all subsequent univariate analyses were conducted by examining differences in content and task, collapsed across spatial boundary (see univariate results below). Planned pairwise comparisons (Bonferroni-corrected for multiple comparisons) were then conducted to examine the relationship in activity between texture and layout for natural and manufactured scenes in each ROI. Since no differences between hemispheres were observed for all ROIs, bilateral regions were combined for analysis.

Figure 9. Regions of Interest. Functionally-defined ROIs are shown on a representative observer’s brain. Talairach coordinates for peak voxels of each ROI in this representative observer are shown as follows: LPPA, -22, -43, -9; RPPA, 31, -45, -5; LRSC, -16, -54, 7; RRSC, 22, -49, 7; LOPA, -29, -84, 15; ROPA, 40, -83, 17; LLO, -39, -72, -7; RLO: 45, -69, 2; LFFA, -36, -34, -23; RFFA, 45, -45, -12; LEVC, -6, -87, -3; REVC, 10, -84, -4.
**Multivoxel Pattern Analysis (MVPA)**

Consistent with Chapter 2, two complementary types of classification analyses were performed with the SVMs; one in which the multiple pairwise results were combined in order to produce multiclass discriminations (distinguishing among all 8 of the condition types) and another in which the individual pairwise discriminations were examined and tested separately. If multiclass decoding is at chance levels, then classification performance will be at \( \frac{1}{8} = 12.5\% \). For all multiclass discriminations, decoding significance was statistically assessed across observers (for each ROI and condition epoch) using one-tailed t-tests versus 12.5\% chance decoding. In contrast, the pairwise discrimination approach allowed us to identify ROIs encoding scene content and spatial boundary while other scene attributes were held constant, as well as make comparisons across tasks (i.e., texture and layout). It is important to recognize that this pairwise information and any nuances in the pattern of effects would be largely obscured using a multiclass discrimination approach. For pairwise discriminations, decoding significance was statistically assessed across observers using one-tailed t-tests versus 50\% chance decoding. Importantly, for both the multiclass and pairwise discriminations, an FDR correction of \( q \leq 0.05 \) was applied based on the number of ROIs examined (Benjamini and Hochberg, 1995).

**Behavioural data analysis**

Behavioural performance measures of accuracy were recorded using E-Prime 2.0 software and analyzed with SPSS, by performing a 2 (spatial boundary: open vs. closed) x 2 (content: natural vs. manufactured) x 2 (task: texture vs. layout) repeated-measures ANOVA.

**Results and Discussion**

**Univariate analysis**

An initial repeated-measures ANOVA including ROI as a factor (PPA, RSC, OPA, LO, FFA, and EVC) revealed significant main effects of ROI \( (F_{5,30} = 21.01, p < 0.001) \), content \( (F_{1,6} = 10.43, p = 0.018) \), and task \( (F_{1,6} = 133.57, p < 0.001) \), but not spatial boundary \( (F_{1,6} = 0.22, p = 0.66) \). Additionally, significant interactions between ROI and content \( (F_{5,30} = 12.30, p < 0.001) \)
and task \((F_{5,30} = 4.95, p = 0.002)\), were observed, demonstrating functional differences in scene-, object-, face-selective, and early visual cortex for the task and stimuli. Thus, all subsequent analyses on each ROI were conducted individually. Moreover, since no main effect of spatial boundary was observed (and no two-way interactions with this factor and ROI), this investigation collapsed across spatial boundary in subsequent analyses to examine differences between content and task (Figure 10).

Further examination of scene-selective cortex revealed a significant main effect of content (PPA: \(F_{1,11} = 58.20, p < 0.001\); RSC: \(F_{1,10} = 44.29, p < 0.001\); OPA: \(F_{1,8} = 6.92, p = 0.030\)), and task (PPA: \(F_{1,11} = 13.62, p = 0.004\); RSC: \(F_{1,10} = 8.73, p = 0.014\); OPA: \(F_{1,8} = 11.43, p = 0.010\)), reflecting higher activity when observers attended to manufactured, over natural, scenes, and layout over texture, respectively. A significant interaction between content and task was found in PPA (\(F_{1,11} = 8.55, p = 0.014\) and RSC (\(F_{1,10} = 9.19, p = 0.013\), but not OPA (\(F_{1,8} = 3.43, p = 0.101\)). When examining the object-selective LO, a significant main effect of task (\(F_{1,11} = 9.03, p = 0.012\), but not content (\(F_{1,11} = 0.05, p = 0.82\) was observed (and no significant interaction), revealing higher activity when observers attended to layout over texture, with no difference between manufactured and natural scenes. In contrast, when examining activity in both EVC and FFA, no significant main effects or interactions were observed (all \(Fs < 2.39\); all \(ps > 0.15\)), indicating no differentiation of scene content or task in these ROIs.

**Figure 10. Univariate Results.** BOLD signal activation for natural (N) and manufactured (M) scenes when attending to either layout or texture. **\(p < .01\), ***\(p < .001\)
Profiles of neural activity for each task (i.e., texture and layout) were consistent among scene-selective regions, revealing no significant region-by-task interactions (all $F_s < .978$; all $p_s > 0.35$). Interestingly, a region-by-content interaction was observed between PPA and OPA ($F_{1,8} = 7.09, p = 0.029$), but not RSC and OPA ($F_{1,7} = 2.89, p = 0.133$). Conversely, profiles of activation in scene-selective cortex were significantly different from those observed in object-selective cortex (significant region-by-content interaction between LO and PPA, RSC, and OPA; all $F_s > 13.20$; all $p_s < 0.01$), face-selective cortex (significant region-by-content interaction between FFA and PPA, RSC, and OPA; significant region-by-task interaction between FFA and PPA, OPA; all $F_s > 15.41$; all $p_s < 0.01$), and EVC (significant region-by-content interaction between EVC and PPA, and RSC; all $F_s > 22.76$; all $p_s < 0.002$; significant region-by-task interaction between EVC and PPA, and OPA; all $F_s > 8.99$; all $p_s < 0.020$). This demonstrates that the results are distinct to high-level scene-selective visual cortex, and are not likely explained by appealing to differences in low-level visual features across scene categories.

To test the hypothesis that neural activity in scene-selective cortex would be modulated by the diagnostic relevance of scene features, planned pairwise comparisons were conducted to examine the relationship between scene content (natural vs. manufactured) and task (texture vs. layout). In line with my predictions, the results demonstrated equal sensitivity to texture and layout in natural scenes (PPA: $t_{11} = 0.20, p = 0.84$, RSC: $t_{10} = 0.43, p = 0.67$; OPA: $t_{8} = 0.77, p = 0.47$), but less sensitivity to texture in manufactured scenes (PPA: $t_{11} = 4.08, p = 0.002$, RSC: $t_{10} = 3.63, p = 0.005$; OPA: $t_{8} = 7.29, p < 0.001$), revealing that the importance of specific scene features (i.e., texture or layout) varies according to the perceived scene content (i.e., natural or manufactured). Notably, no differences were observed in LO, FFA, and EVC (all $t_s < 1.98$; all $p_s > 0.07$).

**Multivoxel pattern analysis**

Previous research has demonstrated that both the content and spatial boundary of a scene can be decoded from scene-selective regions of cortex (e.g., Park et al., 2011), yet less is known about the neural mechanisms underlying these processes. While there is support that edge-based structural information is sufficient for decoding high-level scene content (Walther et al., 2011), the contributions of edge-based information in defining the spatial boundary of a scene, or how
surface characteristics contribute to either of these scene attributes, is unclear. To address these questions, a number of different multivariate analyses were conducted. Following previous investigations (Walther et al., 2009; Park et al., 2011; Kravitz et al., 2011), this analysis first extracted multivoxel fMRI activity and used linear SVM classifiers in each region to examine the extent to which each scene condition could be decoded (Figure 11A). These multiclass discriminations revealed classification accuracies that were significantly above chance (12.5%) for all ROIs (all ts > 3.39, all ps < 0.01), replicating previous findings (see Walther et al., 2009; 2011; Park et al., 2011). To investigate these significant multiclass discriminations in greater detail, subsequent analyses were conducted to examine interactions of task context with high-level scene attributes within each scene category. The classification of task context itself and the distribution of classifier confusion errors across regions of scene-selective cortex were also examined. The results of each analysis are described in turn below.
While multiclass discriminations allow us to investigate decoding of individual scene categories, it does not provide information about the nature of high-level scene attributes within a particular scene category. Thus as a next step in the multivariate analyses, pairwise discriminations were conducted to examine classification accuracy when decoding scene content (train the classifier on the difference between natural vs. manufactured scenes, and test on the same difference, separately in open and closed scenes, i.e., with spatial boundary held constant; Figure 11B, 11D) and spatial boundary (with scene content held constant; Figure 11C, 11E) within a given scene category, separately when observers attended to either layout or texture.

Figure 11. Multivariate Results. (A) Classification accuracy (chance = 12.5%; dashed line) of all eight conditions for each ROI. (B) Decoding accuracy (chance = 50%) of scene content with spatial boundary held constant when attending to layout, and (C) spatial boundary with scene content held constant when attending to layout. (D) Decoding accuracy of scene content with spatial boundary held constant when attending to texture, and (E) spatial boundary with scene content held constant when attending to texture. Error bars represent the standard error of the mean. *p = <.05, **p = <.01, ***p = <.001, +q ≤ .05.
Examination of scene content while spatial boundary was held constant revealed classification accuracy significantly above chance in PPA for content in both open and closed scenes when observers attended to both layout (Open: $t_{11} = 4.89, p < 0.001$; Closed: $t_{11} = 3.99, p = 0.002$; Figure 11B) and texture (Open: $t_{11} = 2.29, p = 0.043$; Closed: $t_{11} = 5.62, p = 0.002$; Figure 11D).

Similar results were found in OPA for both the attend layout condition (Open: $t_{8} = 5.27, p < 0.001$; Closed: $t_{8} = 4.11, p = 0.003$) and the attend texture condition (Open: $t_{8} = 4.97, p = 0.001$; Closed: $t_{8} = 2.99, p = 0.017$), and RSC for both the attend layout condition (but note that while RSC showed significant decoding of content in closed scenes, $t_{10} = 3.91, p = 0.003$, this result was only marginally significant in open scenes, $t_{10} = 1.98, p = 0.076$) and the attend texture condition (Open: $t_{10} = 2.28, p = 0.046$; Closed: $t_{10} = 5.38, p = 0.003$). While classification accuracy in EVC was significantly above chance for scene content in both open and closed scenes in the layout condition (Open: $t_{10} = 2.93, p = 0.015$; Closed: $t_{10} = 3.78, p = 0.004$), it was significant in only closed scenes in the texture condition (Open: $t_{10} = 1.47, p = 0.17$; Closed: $t_{10} = 4.02, p = 0.002$), further dissociating activity in areas of early visual cortex with scene-selective regions.

Following previous research using similar stimulus sets (Park et al., 2011; Kravitz et al., 2011), this investigation next examined classification accuracy of spatial boundary (train and test on open vs. closed scenes) while scene content was held constant (either within natural or within manufactured scenes). Specifically, it was examined whether significant decoding of spatial scene features from scene-selective cortex could occur in manufactured, but not natural, scenes, when observers attended to layout, which would be consistent with the initial hypothesis and univariate results. Indeed, classification of spatial boundary in PPA and RSC were significantly above chance for manufactured (PPA: $t_{11} = 2.43, p = 0.034$; RSC: $t_{10} = 3.35, p = 0.007$), but not natural, (PPA: $t_{11} = 1.17, p = 0.27$; RSC: $t_{10} = 0.81, p = 0.44$; Figure 11C), scenes. Patterns of classification in OPA (Natural: $t_{8} = 3.17, p = 0.013$; Manufactured: $t_{8} = 2.26, p = 0.054$), however, were more similar to EVC (Natural: $t_{10} = 2.61, p = 0.026$; Manufactured: $t_{10} = 3.40, p = 0.007$). Interestingly, unlike the results observed for scene content, when classifying spatial boundary, quite different results were observed when observers attended to the texture, compared with the layout, of a scene (i.e., compare Figures 11B vs. 11D with Figures 11C vs. 11E). Specifically, these results showed significantly above-chance classification accuracy of spatial
boundary in PPA, OPA, and EVC for both natural (PPA: $t_{11} = 3.91, p = 0.002$; OPA: $t_{8} = 3.15, p = 0.014$; EVC: $t_{10} = 4.17, p = 0.002$) and manufactured (PPA: $t_{11} = 2.87, p = 0.015$; OPA: $t_{8} = 3.27, p = 0.011$; EVC: $t_{10} = 2.74, p = 0.021$) scenes, but only for natural scenes in RSC (natural: $t_{10} = 3.22, p = 0.009$; manufactured: $t_{10} = 1.81, p = 0.10$).

**Classification of task context**

In Chapter 2, results indicated sensitivity in scene-selective cortex to both geometry and surface properties. Similarly, previous reports have indicated sensitivity to processing both texture (e.g., Cant and Goodale, 2007; 2011) and spatial layout (e.g., Epstein and Kanwisher, 1998, Epstein et al., 2003) information in scene-selective cortex. Thus, this investigation next examined whether task context (attend texture vs. layout) could be decoded from areas of scene-selective cortex (PPA, RSC, OPA) across each of the four scene categories (see Figure 8), which may ultimately speak to whether their processing is mediated by shared or distinct neural mechanisms. Critically, no significant decoding was found between these tasks in both PPA and RSC (all $t$s < 2.44, all $p$s > 0.13), suggesting similar underlying neural representations between the processing of scene texture and layout in these regions (i.e., the patterns of activation for these two attended scene features were quite similar in PPA and RSC) consistent with the results of Chapter 2. Interestingly, significant decoding of layout versus texture was observed in OPA (Natural Open: $t_{8} = 4.00, p = 0.016$; Manufactured Closed: $t_{8} = 3.27, p = 0.046$), replicating the results of Chapter 2. The similarities between the PPA and RSC are consistent with the univariate results, and together with significant decoding in the OPA, may speak to a functional dissociation between these regions that are based on differences in processing low-level visual information. However, caution must be applied both when interpreting null effects with multivoxel data (see Dubois et al., 2015) and when interpreting apparent functional dissociations across cortical regions.

**Classification Errors**

To further investigate the underlying structure of representations across regions and conditions, the investigation next examined, based on the results of the multiclass
discriminations, the distribution of classifier guesses via a confusion matrix. In a confusion matrix, each row indicates instances of the actual trial class and each column indicates the trial class predicted by the trained SVM classifier. Thus the confusion matrix provides not just a visualization of the correct classifications (indicated by classifier responses along the diagonal axis) but also the cases of misclassification (i.e., where the trained classifier ‘confuses’ the actual trial class with that of another class, as indicated by the off-diagonal classifier guesses). The distribution of misclassifications can be informative as it suggests similarity in the patterns of activity across trial types (i.e., two conditions represented similarly are more likely to be misclassified as one another), which is not necessarily evident from the multiclass decoding accuracies alone.

To extend upon previous work using a similar analysis (Park et al., 2011), here, the investigation determined the types of classification errors made across scene categories while observers attended to individual scene features (texture vs. layout), rather than the image as a whole (as has been done previously). Importantly, this level of analysis allows greater specification of the types of classifier errors made when attending to different features within a scene, and thus greatly improves the understanding of how task demands shape visual representations in scene-selective cortex. If a given region is sensitive to the spatial boundary of a scene when observers attended to either scene layout or texture, then classifier errors in the confusion matrix may be grouped by spatial boundary, regardless of scene content (e.g., open natural scenes may be confused with open manufactured scenes, and closed natural scenes may be confused with closed manufactured scenes, but it is less likely that open and closed scenes would be confused for each other; see hypothetical confusion matrices in Figure 12A). Conversely, if a given region is sensitive to the content of a scene when observers attended to either scene layout or texture, then confusion errors may be grouped by scene content, regardless of spatial boundary (e.g., open natural scenes may be confused with closed natural scenes, and open manufactured scenes may be confused with closed manufactured scenes, but it is less likely that natural and manufactured scenes would be confused for each other).
Figure 12. Confusion matrices generated from the multiclass discriminations. (A) Hypothetical confusion errors within the same Spatial Boundary or within the same Content. Note that when decoding is perfect, the confusion matrix will have a diagonal containing values of 1 and the rest of the matrix will be zero. (B) Confusion matrices for each ROI representing classifier errors across conditions. The average classifier response proportions across observers are shown. Shaded squares represent predicted decoding values significantly greater than 12.5% (chance). (C) Confusion errors (collapsed across task) within the same spatial boundary or within the same content across ROIs. Error bars represent the standard error of the mean. N = Natural, M = Manufactured, O = Open, C = Closed, *p < .05, **p < .01.
Since distinctive structural features across manufactured and natural environments may be used to accurately discriminate scene content (Walther et al., 2011), a greater number of confusion errors may be observed within the same content, rather than within the same spatial boundary when observers attended to scene layout. For instance, when attending to the layout of a scene, manufactured environments may be confused with each other, but not with natural environments, due to characteristic structural components across manufactured scenes (Oliva and Torralba, 2001), regardless of spatial boundary, which may be of greater importance for distinguishing scenes based on depth. Similarly, spatial boundary may be less important when discriminating texture across scenes. Thus, it was hypothesised that overall greater confusion errors in the PPA within the same content would be found, compared with the same spatial boundary, when observers attended to both the layout and texture of a scene.

Figure 12B shows the confusion matrices for each ROI, and Figure 12C shows the types of classification errors made across all ROIs. A repeated-measures ANOVA with factors ROI, scene attribute (spatial boundary vs. content), and task (texture vs. layout) only revealed a significant ROI-by-attribute interaction ($F_{5,30} = 3.01, p = 0.025$), warranting further investigation of differences between spatial boundary and scene content across ROIs. However, since no main effect of task was observed ($F_{1,6} = 2.60, p = 0.16$) and no interactions with this factor (ROI-by-task, task-by-attribute, ROI-by-task-by-attribute; all $F$s < 2.10, all $p$s > 0.093), this investigation collapsed across task for subsequent analyses. Pairwise comparisons were performed using a two-tailed t-test. As expected, confusions within the same content were found to be significantly higher than confusions within the same spatial boundary in PPA ($t_{11} = 3.29, p = 0.007$). Interestingly, RSC showed this same distribution of errors ($t_{10} = 2.29, p = 0.045$), building on both the univariate and multivariate findings showing evidence for similarities in representations between PPA and RSC. In contrast, other regions showed no such differences (all $t$s < 0.89, all $p$s > 0.40) in the types of classification errors made. Critically, and in line with the univariate and multivariate results, patterns of classification errors in PPA were dissociated from those in LO, FFA and EVC (all $F$s < 1.53, all $p$s > 0.26), suggesting the patterns of misclassifications observed in PPA were specific to high-level scene processing and were not likely driven by low-level image properties.
Contrary to previous work (Park et al., 2011), analysis of the distribution of classification errors across scene conditions via the confusion matrix revealed greater clustering of errors within the same content in PPA and RSC, compared with the same spatial boundary. Moreover, these results further differ from this previous work in the finding that the pattern of misclassifications in PPA and RSC were dissociable from the pattern observed in early visual cortex. These findings suggest that the structure of encoding observed in PPA extends beyond the low-level properties of a scene, and reflects the high-level representations of particular features within a scene. For instance, misclassifications clustered within the same content may be indicative of an increased reliance on cues provided by relatively stable structural differences across scene categories. These cues may support the discrimination of features such as layout and texture, whereas spatial boundary attributes may be less relevant for distinguishing these features, and instead may be more relevant for determining spatial depth and routes for navigation.

How can I reconcile these findings with those from previous work? The task was unique as it examined attention to specific features within a scene (i.e., layout vs. texture), and as such, I propose the differences between these findings and those from other researchers stem from the task demands and goals of the observer (Harel et al., 2014), the diagnostic relevance of different features across scene categories and the differential allocation of attention to these features, and differences across the stimuli themselves. For instance, to control for interference effects of salient objects (Davenport and Potter, 2004; Joubert et al., 2007; Lowe et al., 2015), the scenes used in this investigation were selected to be devoid of foreground objects, which would likely explain differences in the representations of information in object-selective cortex when compared with previous work which found greater confusions of content in LO (Park et al., 2011). Indeed, across analyses, activity patterns in LO were dissociated from those in scene-selective cortex, and did not show differences in confusion errors distributed across high-level scene attributes. Moreover, the results from Chapter 2 provide evidence for an effect of object-scene interactivity, and therefore it is highly likely that the presence, or absence, of salient object information may have influenced these results.
**Behavioural performance**

To ensure attention during the experimental task, observers were asked to compare the identity of two images along a particular relevant feature (i.e., layout or texture). The overall response accuracy was high ($M = 95.45 \pm 3.66\%$, range = 90.67 – 97.27 %), confirming observers attended to the relevant feature successfully. Behavioural results are reported in Table 2 (results for accuracy are reported, and not response time, since the former was emphasized, but not the latter, when describing the task to observers; thus interpreting results for accuracy is more valid in this study). Although near-ceiling performance across all conditions was observed, significant main effects of content ($F_{1,11} = 8.17, p = 0.016$), and task ($F_{1,11} = 5.46, p = 0.039$), but not spatial boundary ($F_{1,11} = 3.39, p = 0.093$) were found. Multiple two-way interactions (content-by-boundary: $F_{1,11} = 8.08, p = 0.016$; content-by-task: $F_{1,11} = 19.50, p = 0.001$; boundary-by-task: $F_{1,11} = 8.28, p = 0.015$) were also observed. These significant results are likely driven by comparatively lower performance in one condition, but despite this, behavioural response profiles did not match fMRI response profiles in scene-selective cortex (e.g., in manufactured scenes, activation to layout was greater than texture, despite no difference in behavioural performance across these conditions). This makes it unlikely that task difficulty directly contributed to the observed neural activation patterns in scene-selective cortex. This is consistent with previous findings showing that modulation of activity in PPA was dissociable from manipulations of task difficulty (Xu et al., 2007).

<table>
<thead>
<tr>
<th></th>
<th>Layout</th>
<th>Texture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manufactured</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Open</td>
<td>95.65 ± 1.11</td>
<td>96.57 ± 0.65</td>
</tr>
<tr>
<td>Manufactured</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Closed</td>
<td>96.92 ± 0.57</td>
<td>96.00 ± 1.30</td>
</tr>
<tr>
<td>Natural</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Open</td>
<td>97.27 ± 0.58</td>
<td>94.52 ± 0.96</td>
</tr>
<tr>
<td>Natural</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Closed</td>
<td>95.96 ± 0.68</td>
<td>90.67 ± 1.20</td>
</tr>
</tbody>
</table>

All values represent mean (percent correct) ± SE.
Summary

The results reported in this chapter support the findings of the previous chapter, which have shown that scene-selective cortex represents multiple visual features, such as layout and texture. Extending these findings, Chapter 3 shows that the representations of these features are shaped by perceived scene category, and that these results are dissociated from activity in early visual cortex and areas nonselective to scene processing. Furthermore, differences in the multivariate patterns of activation observed across regions of scene-selective cortex suggest differentiation in the representations of visual features and scene attributes across the broader scene-processing network. The present chapter therefore ties together multiple factors (high-level scene attributes, task context, and individual visual features), which were not jointly investigated previously, to highlight the fact that scene perception and recognition, and visual processing more generally, are performed to serve a particular goal, and it is this goal which determines how informative a particular visual feature is in a particular environment.
Chapter 4

Temporal Dynamics Underlying Scene Categorization in the Human Brain

In previous chapters, this dissertation examined how scene-selective cortex represents visual feature information within a scene, and how this information is integrated with task context and higher-order properties to form a flexible and holistic percept of our local visual environment. In Chapter 2, across three experiments, equal weighting was observed for geometry and surface properties in representations of scene perception, and these features could not be discriminated from one another using multivoxel pattern analysis, suggesting shared underlying neuronal mechanisms. These findings were replicated in Chapter 3, where it was again found that these features could not be discriminated from patterns of activity within the scene-selective PPA, yet attending to these features separately influenced neural categorization of scenes, supporting a role for task context and feature diagnosticity in shaping scene representations. In addition to these findings, previous investigations have revealed neural categorization accuracy of scenes was found to be similar between colour photographs and line drawings of scenes (Walther et al., 2011). Together with the findings in previous chapters, this suggests scene geometry and surface information may be processed similarly within the spatial domain.

The neural mechanisms underlying the processing of these features may therefore be overlapping within the spatial domain, affording us limited precision to explore their individual contributions. Yet there is evidence to suggest that edge-based information is processed with higher temporal priority than surface information within the visual stream (Fu et al., 2016), supporting the notion that these features may be dissociated across time. Within the temporal domain, their processing may unfold on different timescales along a hierarchy of visual processing. Thus, the aim of the present chapter is to examine how these visual features influence neural categorization of scenes across time. Here, this research will build on previous EEG and MEG work which has provided evidence for the categorization of object information as early as 150ms (Thorpe, Fize, and Marlot, 1996; Vanrullen and Thorpe, 2001), and scene information within 250ms (Groen et al., 2013; Cichy et al, 2016; Harel et al., 2016). This chapter will investigate both basic-level scene categories, which correspond to the most common
categorical representation (e.g., forest, mountain), as well as global scene properties, which correspond to various kinds of abstraction, and include descriptors of scene content (natural or manmade) as well as spatial boundary (open or closed scene boundaries) (Oliva and Torralba, 2006). Behavioural evidence has demonstrated that basic categories emerge prior to the categorization of global properties (Rosch et al., 1976; Tversky and Hemenway, 1983), yet there is growing evidence to suggest the opposite is true (Greene and Oliva, 2009; Loschky and Larson, 2010; Kadar and Ben-Shahar, 2012; Sun et al., 2016). Therefore, in addition to exploring how line drawings and colour photographs influence the time course of scene categorization, this chapter will investigate the earliest points in time at which different levels of categorization emerge.

To do so, the present study applied a multifaceted approach using EEG to examine the neural time course of natural scene processing for colour photographs and line drawings of six scene categories (beach; city; forest; highway; mountain; office). This approach explored the temporal dynamics of scene perception across both basic category levels and global scene properties using not only traditional event-related potentials (ERPs), but also a novel classification analysis which examined the correlations of within-category versus between-category discriminations at a precise temporal scale. Critically, this investigation adopts an orthogonal memorization task to explore the earliest time points at which scene categories can be decoded from electrophysiological activity, thus allowing examination of the neural signatures of scene categories independent of direct behavioural goals.

**Method**

**Observers**

Sixteen paid observers (13 females, mean age 19.1 ± 2.3 years; all right-handed) with normal or corrected-to-normal visual acuity and no history of neurological impairments were recruited from the University of Toronto community. This sample size is consistent with those used in previous research exploring image categorization with EEG (e.g., Thorpe, 1996; Vanrullen and Thorpe, 2001; Groen et al., 2013; Harel et al., 2016). All observers gave informed
consent in accordance with the University of Toronto Ethics Review Board. No observers were excluded from analysis.

**Stimuli and Procedure**

The experiment was programmed, displayed, and analyzed using Matlab (MathWorks, Natick, MA) software running on a desktop computer, with a ViewSonic 21-inch monitor (1280x1024 resolution, 85 Hz refresh rate). The viewing distance was 57 cm, and observers made responses with their right hand on the mouse, and their left hand on the “z”, “x” and “enter” keys of a standard keyboard. Continuous, unreferenced EEG was recorded at a sampling rate of 512 Hz using a BioSemi ActiveTwo system with 64 Ag/AgCl scalp electrodes in standard 10-20 placement with additional electrodes at each mastoid for off-line re-referencing, below each eye for blink detection, and at the outer canthus of each eye for lateral eye movement detection. All re-referencing and filtering was done off-line using ERPLAB software.

Four-hundred thirty two colour photographs of six real-world scene categories (three natural: beaches, forests, and mountains; three man-made: city streets, highways, and offices; 72 per category; See Figure 13 for examples) were used following previous research (Walther et al., 2011). These images were chosen from a set of 4,025 images downloaded from the Internet as the best exemplars of their categories according to ratings by an average of 137 observers per image (Torralbo et al., 2013). Images were resized to 800 × 600 pixels. Line drawings were produced by trained artists at the Lotus Hill Research Institute (Wuhan, Hubei Province, People’s Republic of China), who traced outlines in the colour photographs using a custom graphical user interface. Line drawings were rendered by connecting the anchor points with black straight lines on an 800- × 600-pixel white background.
Observers were instructed to view and memorize a series of images, after which they would receive a recall test. During each trial, observers were asked to maintain central fixation and refrain from all eye movements, including blinking. In the study-phase, images were presented within twelve randomized blocks. Each block contained sixty-six randomly-presented colour photographs or line drawings from any of the six scene categories. Images from each stimulus category were evenly distributed across all blocks. After the initial instruction to view and memorize a series of images, observers initiated each block by pressing “enter” on the keyboard, and began each trial by clicking the left mouse button. Trials were self-initiated to allow for sufficient time for rest between trials which would therefore minimize artifacts caused by eye movements. The onset of each trial with respect to observers’ button presses was randomly jittered across three temporal selections (300ms, 500ms, or 700ms), and each image was displayed for 1500ms. Following the presentation of sixty-six randomly presented study images, observers were asked to recall whether an image was new or old during the test-phase. The test-phase consisted of twelve self-initiated and randomly-presented trials containing six new and six old images, and observers were instructed to press the “x” key if the image was old (familiar), and the “z” key if the image was new (unfamiliar). Test-phase images remained on screen until response.
Analysis

Each individual data set was filtered using a finite impulse response (FIR) filter high-passed at 0.01 Hz and low-passed at 70 Hz. The continuous data were then re-referenced to the average of the two mastoids. Lateral eye movements were detected using a step-like function (horizontal eye electrodes, threshold: 80 μv) and blinks were detected using a moving window peak-to-peak threshold (vertical eye electrodes, threshold: 30 μv). Data was segmented into 1200ms analysis windows – composed of the 200ms immediately preceding onset of scene stimuli as the baseline period, which was baseline corrected to zero, and 1000ms post-stimulus onset as the critical window. Analyses focused on three electrode sites: centro-occipital (Oz), right centro-parietal (P4), and right centro-frontal (F4).

To identify the time-course of scene content and boundary activity, $t$-statistics were calculated comparing the amplitude of pairs of ERPs (e.g., natural versus man made) at each time point of the ERP, using an alpha value of 0.01, to determine the first time point at which they differed. Given that this entailed calculating a large number of $t$-values, the possibility of a false alarm is high. To account for this possibility, this investigation sought to distinguish between spuriously significant comparisons at random points in the time series from meaningfully significant comparisons that would cluster in time (that is, consecutive time-points would also show differences in amplitude). To this end, the number of consecutive, significant time points that would be expected by chance in a series of paired comparisons where no differences exist were calculated. In this case, any given time point has a well-defined probability of reaching significance by chance ($p = .01$, given the alpha level used here). The probability of a run of significant comparisons of length $k$ using a Monte Carlo simulation with 10,000 iterations was estimated. This simulation revealed that a run of length 10 or greater occurred in fewer than 5% of the simulation runs, and thus I used this as a criterion for considering a given pairwise comparison’s statistical significance as meaningful (i.e., it needed to be followed by at least 9 statistically significant time-points).
Results and Discussion

Global scene properties

Previous research has shown that the distinction between global scene properties, such as the content of a scene (i.e., natural versus manmade), and the spatial boundary, or layout, of a scene (i.e., open versus closed) is an important factor mediating scene perception and our ability to navigate through an environment (Kravitz et al., 2011; Park et al., 2011). Therefore, this analysis began by investigating the time course of these scene properties across scene categories and stimulus type. To do so, scene content was examined by comparing natural (beach, forest, mountain) and manmade (city, highway, office) scenes, and boundary was examined by comparing open (beach, highway) and closed (forest, city) scenes, following previous research (Oliva and Torralba, 2001; 2006). I then examined, in content and boundary separately, whether conditions (e.g., natural versus manmade) could be dissociated from patterns of responses during early electrophysiological activity. By including both colour photographs and line drawings, this investigation could examine the extent to which surface information, such as colour and texture, influences the time course of scene categorization in the human brain. In addition to an examination of early occipital visual cortex (Oz), this investigation also included right-hemisphere parietal (P4) and frontal sites (F4) of interest based in part on previous evidence indicating larger right-hemisphere neural activations to scenes (e.g., Epstein and Kanwisher, 1998) and in part on the distribution of electrophysiological activity across the scalp for all study-phase images averaged across observers, irrespective of image type and scene category (Figure 14).
Figure 14. Site selection (above) was based on overall higher right-hemisphere response amplitudes using scalp maps averaged across observers for colour photographs and line drawings (below) from 100-800ms. Sites included a centro-occipital site (Oz), a right centro-parietal site (P4), and a right centro-frontal site (F4), indicated in red.

An ERP analysis was performed for both colour photographs and line drawings, separately, for scene content (Figure 15), and scene boundary (Figure 16). After computing a difference wave for conditions in each of the sites of interest separately, this analysis then compared each time point in the difference waves to zero using a paired-samples $t$-test ($\alpha = .01$) to find the earliest point at which the ERPs diverged. This analysis was therefore able to identify the earliest time point at which a cluster was significantly different from zero using a moving window with a minimum of 10 consecutive significant $t$-tests. The leading edge of said clusters occurred within the early visually-evoked P1 (50 – 130ms) and P2 (150-275ms) components (Busch et al., 2004; Calvo and Beltrán, 2014). Full plots of the ERPs for scene content and scene boundary, with significant temporal clusters highlighted and 95% confidence intervals plotted, can be found in Appendix Item 2 (scene content) and Appendix Item 3 (scene boundary).
Figure 15. Event-related potentials (ERPs) and their difference waves plotted in microvolts (μV) for natural (beach, forest, mountain) versus manmade (city, highway, office) scenes averaged across observers (N = 16) for centro-occipital (Oz), right centro-parietal (P4), and right centro-frontal (F4) sites. An asterisk indicates the earliest significant time point (p < 0.01) within a time component.
When examining scene content in line drawings, the earliest significant discrimination between natural and manmade scenes during the examined time windows occurred during the P1 component (79ms) in the occipital region, the P2 component (178ms) in the parietal region, and the P1 (112ms) and P2 (216ms) components in the frontal region. For colour photographs, the earliest significant discrimination for natural and manmade scenes occurred during the tail-end of the P1 component (130ms), and was continually sustained throughout the P2 component in the occipital region, the P1 (81ms) and P2 (150ms) components in the parietal region, and the P1 (73ms) and P2 (150ms) components in the frontal region. When examining spatial boundary in line drawings, discrimination between open and closed scenes occurred during the P1 component (81ms) in the occipital region, and, interestingly, was absent in both the parietal and frontal regions. For colour photographs, discrimination for open and closed scenes occurred during the P1 (97ms) and P2 (163ms) components in the occipital region, during the P1 component (56ms)
in the parietal region, and during the P1 (52ms) and P2 (169ms) components in the frontal region.

These results show that neural categorization of global scene properties occurs very quickly in both colour photographs and line drawings, even as early as the P1 component. Thus, the findings here provide evidence that neural processing as early as the P1 component (50 – 130ms) with sustained activity extending through the P2 component (150-275ms), may be used to extract and differentiate information from scene categories. Moreover, these results show that global scene properties for line drawings emerge even earlier than colour photographs within occipital cortex, suggesting that the structure preserved in line drawings is sufficient for the extraction of both the content and boundary information of a scene. In contrast to colour photographs, discrimination of these properties for line drawings occurred later, or was entirely absent, over parietal and frontal regions, suggesting information contained in line drawings and colour photographs may each contribute to scene categorization in unique ways, within both the spatial and temporal domains. The structure preserved in line drawings may form a primary and necessary basis for these distinctions, yet a clearer picture may emerge when structure is presented along with colour and texture information.

**Basic-level scene categories**

The next investigation aimed to determine the earliest time point at which basic category levels of scenes can be extracted from neural activity. First, the averaged ERPs were plotted for each scene category for both colour photographs and line drawings (Figure 17). To assess the time-course of scene category selectivity in ERPs, the similarity of voltage changes over time was analyzed at different electrodes within and between scene categories. To do this, two ERPs for each category were computed. For each observer, scene category (beach; forest; mountain; city; highway; office) and stimulus type (colour photographs; line drawings) trials that remained after artifact rejection were randomly divided into two disjoint sets. Twenty four grand-average ERPs were then calculated by averaging across trials and observers for each scene category, stimulus type, and subset. Next, this investigation calculated similarity matrices by calculating the correlation of the grand average ERPs at each of the electrodes of interest for each scene category in set 1 with those in set 2, separately for photographs and line-drawings (Figure 18A).
The diagonal entries of these matrices contain correlations of ERPs for matching categories, the off-diagonal entries for non-matching categories. Scene selectivity was quantified by contrasting the average similarity (i.e., \( r \) value) of the diagonal entries with the average similarity of the off-diagonal entries (Figure 18B). Statistical inferences were made using independent samples \( t \) tests (equal variance not assumed) comparing these same values (i.e., on- vs. off-diagonal entries). In each of these cases, \( r \) values were transformed to a normal distribution using the inverse hyperbolic tangent (Fisher’s \( z \) transform). Three sites of interest were selected on the basis of earlier ERP analyses: centro-occipital (Oz), right centro-parietal (P4), and right centro-frontal (F4).

Figure 17. Event-related potentials (ERPs) plotted in microvolts (\( \mu \)V) for all scene categories (beach, forest, mountain, city, highway, office) averaged across observers (\( N = 16 \)) for centro-occipital (Oz), right centro-parietal (P4), and right centro-frontal (F4) sites.
The similarity of elements was higher on than off the diagonal for both photographs and line-drawings for occipital ($t_s > 7.4$, $p_s < 0.001$), parietal ($t_s > 2.5$, $p_s < 0.05$), and frontal ($t_s > 4.5$, $p_s < 0.001$) sites, providing the first evidence that basic-level categories can be decoded from electrophysiological activity measured from the scalp. Given interest in the time-course of this similarity, however, similarity indices were calculated again, but using a moving window containing 30 samples from ERPs in each window (spanning 58.6ms). Thirty samples were used in this analysis to strike an optimal balance between temporal resolution (i.e., smaller time windows) and statistical power (i.e., reliable estimations of the correlation between time-series). These time courses of scene selectivity are shown in Figure 19. Marked with red asterisks are time points where the similarity is statistically significant at alpha < 0.01. For the occipital region, colour photographs show scene selectivity as early as 100ms, and for line drawings, scene categories can be distinguished as early as 78ms. For frontal and parietal regions, selectivity is much more sporadic, and always later, although it is interesting to note that within
parietal sites, scene category discrimination succeeded within the first 200ms for colour photographs, but not line drawings. Note that, given the moving-window approach, these are conservative estimates: the 100ms estimate of scene category discrimination at Oz for colour photographs, for example, is based on a correlation of the voltage in the 40ms-100ms range window of grand average ERPs, and thus may reflect differences in neural responses even earlier than 100ms for different scene categories. Given that real-world scenes contain a high degree of statistical regularity (Oliva and Torralba, 2006; Torralba and Oliva, 2003), findings within occipital regions are consistent with previous research which has suggested that global scene information can be computed from responses of early visual areas (Groen et al., 2013). In the present findings, image statistics may form a powerful basis for rapid categorization of scenes within early visual areas, contributing to both the perception of global scene properties and basic-level categorization. Parietal sites showed evidence for scene category discrimination in colour photographs, but not line drawings, revealing differences in the representation of visual information, and how this information is utilized for scene categorizations.
**Figure 19.** Grand-averaged discriminations (N = 16) of scene categories (diagonal vs. off-diagonal) for colour photographs (left) and line drawings (right) and their difference waves in centro-occipital (Oz), right centro-parietal (P4) and right centro-frontal (F4) electrode sites. Asterisks (marked in red) indicate significant discriminations (Fisher Z) across scene categories using a moving-window containing 30 samples from ERPs in each window (spanning 58.6ms). Dashed lines represent the earliest time of significant category discrimination at any of the sites.

*p < 0.01

**Correlations across colour photographs and line drawings**

Exemplars of real-world scenes within the same category contain a high degree of statistical regularity (Oliva and Torralba, 2006; Torralba and Oliva, 2003), and their corresponding image statistics may capture even global properties of a scene, such as scene content (Groen et al., 2013). Given both similarities and differences across the time course of scene processing for colour photographs and line drawings, an important question concerns
whether low-level image statistics capture category structure across different image types, or whether these category regularities are driven by higher-order representations and meaning. For instance, if significant correlations of scene categories are absent across stimulus types, this would suggest that low level visual information contained within the stimulus sets may be sufficient to capture unique category structure. Previous investigations have revealed that object categories emerge in neurophysiological activity within 100ms, which may correspond with an initial feedforward perceptual stage reflecting responses of the human visual system to low-level image properties (Vanrullen and Thorpe, 2001). Similarities between scene categories in colour photographs and line drawings were therefore examined, across the sites of interest. First, similarity matrices were created for all scene categories across colour photographs and line drawings (Figure 20). It was then examined whether diagonal entries (e.g., colour photographs of beaches versus line drawings of beaches) would differentiate from off-diagonal entries (e.g., colour photographs of beaches versus line drawings of highways) across the full time course of scene perception, by converting $r$ scores to Fisher Z values and performing an independent samples $t$-test (equal variance not assumed).

**Figure 20.** Similarity matrices showing the grand-averaged correlations of each scene category for all observers ($N = 16$) averaged across time (0 – 1000 ms) in centro-occipital (Oz), right centro-parietal (P4) and right centro-frontal (F4) electrode sites in both colour photographs and line drawings.
The similarity of diagonal elements did not differ from off-diagonal elements in parietal or frontal regions ($t_s < 1.59, p_s > 0.11$), suggesting that the category structure observed for colour photographs and line drawings may be driven by low level properties that do not easily transfer between the two types of images. A significant difference between diagonal and off-diagonal elements at the occipital site was observed ($t_{70} = 2.48, p = 0.016$). Examination of this time course showed a significant category signal starting at 340ms, which is presumably related to feedback from higher-level brain regions following an initial feedforward perceptual encoding stage (Vanrullen and Thorpe, 2001). These higher-level, later-stage signals are likely to encode higher-order scene properties, which, unlike low-level features, transfer between line drawings and colour photographs (Figure 21). Later stages of scene processing may represent the abstract identity of a scene and thus may be tolerant to low-level image changes (Dilks et al., 2011).

![Graph](image)

**Figure 21.** Grand-averaged discriminations ($N = 16$) of scene categories (diagonal vs. off-diagonal) comparing colour photographs and line drawings and their difference waves in the centro-occipital (Oz) region. Asterisks (marked in red) indicate significant discriminations (Fisher Z) across colour photographs and line drawings using a moving-window containing 30 samples from ERPs in each window (spanning 58.6ms). *$p < 0.01$

**Behavioural Accuracy**

Observers performed well on the memorization test with an average accuracy of $68.5\% \pm 15.5\%$, confirming their attention to the task. To further analyze behavioural accuracy for recognition performance during the test-phase, a two (stimulus type: colour photographs versus...
line drawings) by six (category: beach, city, forest, mountain, highway, office) repeated-measures ANOVA was conducted. This analysis revealed main effects of stimulus type ($F_{1,15} = 43.27, p < 0.001$) and category ($F_{5,75} = 7.82, p < 0.001$), but no interaction ($F_{5,75} = 1.83, p < 0.118$). Overall, recognition rates were higher for colour photographs ($M = 73.8\% \pm 14.2\%$) than line drawings ($M = 63.37\% \pm 15.0\%$). Response accuracy for all categories can be seen in Table 3.

<table>
<thead>
<tr>
<th></th>
<th>Beaches</th>
<th>City</th>
<th>Forest</th>
<th>Highway</th>
<th>Mountain</th>
<th>Office</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Colour Photographs</strong></td>
<td>72.40</td>
<td>79.69</td>
<td>73.96</td>
<td>73.44</td>
<td>58.85</td>
<td>85.38</td>
</tr>
<tr>
<td>± 2.82</td>
<td>± 3.22</td>
<td>± 1.84</td>
<td>± 4.18</td>
<td>± 2.69</td>
<td>± 2.93</td>
<td></td>
</tr>
<tr>
<td><strong>Line Drawings</strong></td>
<td>65.15</td>
<td>66.15</td>
<td>57.81</td>
<td>67.19</td>
<td>56.25</td>
<td>67.19</td>
</tr>
<tr>
<td>± 4.09</td>
<td>± 3.99</td>
<td>± 3.91</td>
<td>± 3.69</td>
<td>± 2.99</td>
<td>± 3.18</td>
<td></td>
</tr>
</tbody>
</table>

*All values represent mean (percent correct) \pm SEM*

**Summary**

The results of Chapter 4 show that the temporal separation of scene-related information occurs with incredible speed in the human brain, supporting an efficient neural network able to discern information from the environment in only an instant. Distinctions of global scene properties, such as content and layout, emerge within a similar time frame to basic-level distinctions (e.g., beach; forest), but also diverge depending on the visual information available (e.g., line drawings versus colour photographs). The structure preserved in line drawings may form a primary and necessary basis for categorization of scenes, yet colour information may contribute to scene recognition in later-stages of processing: In contrast to colour photographs, discrimination of scene categories for line drawings occurred later, or was absent, over parietal and frontal regions. These findings highlight the fundamental role of categorization for the
purposes of perceiving and understanding our visual environment, and the joint importance of structure and surface information underlying the neural representations of scenes.
Chapter 5

General Discussion

The purpose of this dissertation was to examine how different visual features influence the underlying neural representations of the visual world, how these properties interact with feedback processes involving observer goals and task context, and whether these visual features diverge across time and influence neural processing within the temporal domain. Investigations in Chapter 2 examined the weighting of geometry and surface properties in driving neural activation in scene-selective regions. Findings from this chapter indicate that the scene-selective PPA responds equally in strength to the surface properties and geometry of a scene, and these features could not be discriminated from one another. Moreover, the PPA responded with greater sensitivity to the surface information of objects over their geometry, suggesting this region may be particularly sensitive to the processing of surface information such as colour and texture. Chapter 3 presented work which aimed to investigate how geometry and surface properties contribute to neural representations of scenes across different scene types and categories by manipulating attention to these features separately to examine the roles of feature diagnosticity and task context in shaping neural representations of scenes in human visual cortex. Evidence revealed in Chapter 3 demonstrates that, relative to geometry, activity in scene-selective cortex showed equal sensitivity to texture in natural scenes, but less sensitivity to texture in manufactured scenes. These findings suggest activity in scene-selective cortex is not only modulated by multiple scene features, but may be dependent on the relevance of various features within a particular type of scene. Multivoxel pattern analysis further revealed that the encoding of high-level scene attributes varies according to scene category and is influenced by task context, suggesting dynamic and flexible scene representations are formed from an interaction between the processing of multiple scene properties and the task goals of the observer. In Chapter 4, findings indicate that scene categories can be decoded from neural responses within a fraction of a second within the human visual system, and further reveal evidence that visual properties within a scene influence this time course. Specifically, line drawings were found to be sufficient for scene categorization, while colour photographs show clearer patterns of activity.
which emerge later in time. In the following discussion, I will address each of these findings in greater detail.

A role for surface information in shaping neural representations of scenes

The findings presented in Chapter 2 indicate that the scene-selective PPA responds equally in strength to the surface properties and geometry of a scene, and these features could not be discriminated from one another using MVPA. In object and scene perception, evidence revealed preferential sensitivity to an object’s shape within the object-selective LO, but the opposite was found to be true for these object properties within the PPA when objects were presented in isolation (Experiment 1). These results suggest that the PPA may be particularly sensitive to processing surface information. An interaction between objects and scenes, however, modulated this relationship within the PPA such that the surface properties and geometry of objects were processed with similar sensitivity (i.e., similar univariate response amplitudes) when objects were presented within scenes (Experiment 2), despite demonstrating distinct underlying neural representations (i.e., multivariate spatial patterns of activation). Furthermore, varying scene geometry and surface properties independently of each other led to equivalent releases from adaptation in the PPA (Experiment 3). Together with the findings that the PPA can discriminate geometry and surface features of objects but not scenes, this evidence suggests that the processing of scene geometry and surface properties are weighted equally and may share similar underlying neuronal mechanisms in the PPA, and these features may jointly contribute to the construction of scene identity. The present findings also provide evidence that the processing of surface features in the PPA may be mediated by domain-specific, rather than domain-general mechanisms, since greater sensitivity in the PPA to the surface properties of scenes over objects was observed, emphasizing the importance of this property for scene perception. Evidence across three experiments highlights how these global scene properties are processed in the broader scene-processing network, albeit in ways which are dissociated from each other, suggesting a complementary relationship between regions in this network for the purposes of perceiving and navigating the world around us.

While extensive evidence has supported a primary role for the PPA in the encoding of spatial information (Epstein and Kanwisher, 1998; Epstein et al., 2003; Kravitz et al., 2011; Park
et al., 2011), there is considerably less research investigating the encoding of surface information within this region. In anticipation of the present neural findings, however, behavioural results have implicated surface characteristics such as colour and texture as being instrumental in mediating early-stage processes responsible for successful scene recognition (Schyns and Oliva, 1994; Oliva and Schyns, 2000; Goffaux et al., 2005; Castelhano and Henderson, 2008). A case of topographical disorientation (wayfinding difficulties) related to landmark agnosia has revealed that geometry and surface properties may interact and jointly contribute to scene perception (Robin et al., 2017). Converging neuroimaging results have shown that the surface and material properties of objects are processed in regions overlapping with scene-selective areas of parahippocampal cortex in humans (Peuskens et al., 2004; Cant and Goodale, 2007; 2011), and demonstrated sensitivity to both object and scene material properties in nonhuman primate visual cortex, including areas selective to scene processing (Kornblith et al., 2013; Goda et al., 2014). Surface properties may play a meaningful role in humans’ ability to distinguish one scene from another through diagnostic cues used for recognition, and may additionally support various affordance-related tasks such as action planning and navigation. In this dissertation, by using strictly-controlled environments which allow for the direct comparison of visual features, it was demonstrated that scene-selective cortex responds just as strongly to distinctive surface properties as it does to geometry. Together with previous findings, these results raise a question of how multiple cues are used to construct the visual world around us across inherently complex and vastly different environments.

A role for task context and feature diagnosticity in shaping neural representations of scenes

One account to explain how multiple cues are utilized in scene perception may involve the role of diagnostic features and goal states (Chapter 3). In fact, an existing behavioural scene recognition framework centers on the notion of feature diagnosticity: the idea that specific visual cues are used for specific types of categorizations and an interaction between task demands and available visual information can explain how different cues are used to recognize scenes (Oliva and Schyns, 1997). Thus, for a complete view of scene understanding, it is necessary to account for both the contributions of diverse scene properties and differing observer goals (Malcolm et al., 2016). Altering diagnostic aspects of a scene may in turn alter place information responsible
for perceived novelty and the subsequent encoding of a scene in memory, consistent with findings suggesting the PPA is involved in encoding novel place information in memory (Epstein et al., 1999). Surface properties may represent the unique identity of a scene in much the same way that geometry does, yet this representation is dependent on context and scene category. For instance, surface properties may provide valuable information for the formation of scene identity within natural environments, such as forests or deserts, but may be less useful for distinguishing among different cityscapes.

Typically, edge-based structural information is sufficient for observers to distinguish between scenes (Biederman and Ju, 1988; Delorme et al., 2000; Walther et al., 2011). In contrast, the results presented in Chapter 3 indicate that when structural information is less informative (e.g., in natural environments), processing in scene-selective cortex may rely on other diagnostic features which uniquely inform scene identity. Indeed, Steeves and colleagues (2004) have shown that, unlike in healthy observers, PPA activation in an individual with profound visual form agnosia (i.e., impairments in processing visual structure) was modulated by the presence of appropriate scene colour when colour was useful for accurate scene identification. Neural activity in scene-selective regions within the ventral visual cortex may therefore also reflect the relative higher-order contributions, or feedback processes, of various scene features beyond spatial layout which contribute to the formation of scene identity.

In a similar vein, Harel et al. (2014) have suggested that top-down signals produced by behavioural goals and observer intent directly impact visual object representations within the ventral visual pathway, supporting the view that cortical activity reflects not only the physical properties of a stimulus, but also the internal state of the observer. In Chapter 3, I provided evidence consistent with this research by showing that task context (i.e., attended scene feature) influences neural activity in scene-selective cortex. Here, global texture cues may provide meaningful information for discriminating scenes containing a high degree of physical similarity, such as desert landscapes, and thus attending to these features may selectively modulate activity in scene-selective cortex accordingly. In addition to providing meaningful high-level identity information, global texture cues have been shown to inform the processing of spatial depth and contours (Torralba and Oliva, 2003), suggesting a convergent and complementary relationship between texture and spatial structure. Here, I would argue that the PPA may not necessarily be
responsible for the selection of diagnostic information, but rather responds in accordance with top-down feedback mechanisms to facilitate scene recognition. To the best of my knowledge, previous work has not revealed how manipulating attention to a particular scene feature has differential effects on activity in scene-selective cortex, depending on the type of scene being viewed. These current results thus emphasize the fundamental importance of examining the modulation of activity in scene-selective cortex as a function of attending to various scene-specific properties and task-based goals. In the real-world, we are not merely passive observers for the purposes of perception, but use attention to filter the influx of visual information in accordance with our goals. That is, we engage in scene-processing for the purposes of interacting with our environment, and based on the nature of the surrounding environment, different visual features will become more, or less, important.

Consistent with previous research suggesting the PPA represents both content and spatial boundary information (Park et al., 2011), the multivariate results of Chapter 3 revealed that regions in scene-selective cortex discriminated between both of these high-level scene attributes. Having confirmed previous findings, this investigation next employed novel analyses to determine whether the decoding of high-level attributes varies according to task context and scene category. An investigation of within-category discriminations revealed significant decoding of spatial boundary and content when attending to both texture and layout, suggesting that, together with layout, texture may form an important basis for defining scene identity. More specifically, this analysis revealed results consistent with univariate findings: significant decoding of a spatial scene feature (i.e., boundary) only occurred in manufactured, but not natural scenes (in the PPA and RSC), when observers attended to the layout of a scene. The decoding of spatial boundary when observers attended to texture, however, differed markedly from when observers attended to layout. Critically, these findings show for the first time that task context directly impacts the representations of high-level scene attributes, and underscores the notion that scene attributes and task demands may modulate activity in scene-selective cortex to varying degrees across different scene categories. Together with the univariate results, these novel findings suggest the importance of scene features may be scene-specific and task-dependent, rather than of equal importance across different scene categories.
In contrast to the above results, both spatial boundary and content could be decoded when attending scene layout regardless of scene type in areas of early visual cortex, consistent with previous research (Park et al., 2011). Real-world scene categories contain a high degree of statistical regularity reflected in distinct low-level features and global spatial frequency distributions across both scene categories and boundaries (Oliva and Torralba, 2001; Torralba and Oliva, 2003). This leaves open the possibility that scenes could be discriminated based on low-level attributes alone. Thus, an alternative explanation for these findings may lie in the sensitivity of the PPA to low-level visual statistical changes, including those elicited by changes in both spatial structure and surface properties (e.g., monocular cues to depth). This view would deem the PPA an area purely devoted to processing the low-level visual statistics of a scene. The PPA, and scene-selective cortex more generally, is sensitive to processing low-level image features such as high spatial frequency content (Rajimehr et al., 2011; Watson et al., 2016). This may help to explain how regions distributed across scene-selective cortex show similar response amplitudes to changing features (both spatial and non-spatial). Although it is highly likely that low-level features and statistical regularities across scene categories contribute to the activation patterns associated with scenes in the PPA, these neural patterns are unlikely to be driven purely by such features. For example, given findings that (1) the PPA plays a more direct role in humans’ ability to categorize real-world scenes compared with areas of early visual cortex (Walther et al., 2009; 2011), (2) the PPA exhibits higher activation for object-texture over object-shape and vice-versa in LO (Chapter 2), and (3) the functional dissociations between the PPA and control regions (FFA and OFA; Chapter 3) and between the PPA and EVC (Chapter 3), I believe that appealing to the processing of low-level features alone is unlikely to explain the present series of results.

The broader scene-processing network

The results presented in this dissertation have revealed functional dissociations within the scene-processing network, warranting consideration of the role that these ventral regions play in visual perception. Recent investigations have demonstrated the causal involvement of the OPA in boundary perception (Julian et al., 2016), and have suggested that this region may be involved in representing the local elements of a scene, and first-person perspective for visually-guided
navigation (Kamps et al., 2016a; 2016b). In Chapter 2, there is evidence for dissociations between the PPA and the OPA across experiments: the OPA showed greater sensitivity to the processing of scene geometry (Experiment 1), and was able to discriminate scene geometry from surface properties across the first two experiments. Together, these results suggest that scene geometry and texture are processed distinctly within the OPA, and that this region may be particularly sensitive to the processing of scene geometry. The dissociable processing of these features with the PPA may reflect the involvement of the OPA in representing geometry for the purposes of local boundary perception, but further research is necessary to test these ideas directly.

While the results presented in Chapter 2 indicate that RSC was more similar to the PPA overall, some dissociations and interactions across these regions highlight their differences. For example, across experiments, while both regions could not discriminate scene features, an inverse relationship was observed in object perception wherein the PPA was able to discriminate objects features, yet this was not the case in RSC. How these results translate to the involvement of these regions in scene perception, recognition, and navigation warrant further investigation. Moreover, these results highlight the differences across these regions in feature-processing, and provide stronger evidence for the role of the PPA in processing surface properties: While the PPA may be more sensitive to processing the surface properties of an object over its shape, RSC does not show this sensitivity, and may represent these features more generally. These latter findings could be explained by the central involvement of RSC in spatial navigation: RSC has been shown to represent position and orientation relative to fixed elements of the visual world, and may generalize across local environments with similar geometric structures (Marchette, Vass, Ryan, and Epstein, 2014).

Similarly, in Chapter 3, results indicate both similarities and differences in functional representations of scenes across the broader scene-processing network. These regions exhibited similar patterns of univariate activation when observers attended to either scene texture or layout, suggesting reliable and distinguishable representations of diagnostic scene features across the scene-processing network compared with non-scene selective regions. Conversely, the representation of high-level scene attributes (content and spatial boundary) differed across these regions. Across analyses, similar patterns of activation were observed in the PPA and RSC, yet
these patterns were often dissociated from those observed in the OPA. Damage to the PPA and RSC has been associated with deficits in the simple visual identification of scenes or landmarks (Aguirre and D’Esposito 1999; Mendez and Cherrier 2003), and with difficulty using landmarks to orient oneself to navigate through an environment (Takahashi et al. 1997), respectively. While recent work has demonstrated a causal involvement for OPA in scene perception (Dilks et al. 2013), less is known about how different scene properties are represented within this region.

The functional differences observed across the PPA, RSC, and the OPA may be related to differences in how task context and the importance of perceptual features shape representations across areas of scene-selective cortex. In other words, the goals of the observer may differentially affect activation across these regions, by interacting with the type of perceptual information processed within these regions (e.g., both spatial and non-spatial visual information in the PPA, but more weighting towards spatial information in the OPA; Cant and Xu, 2012; 2015). An important future question concerns whether the OPA represents high-level information consistent with the PPA and RSC, or more basic perceptual information which may complement representations in the PPA and RSC. Research has sought to explore these questions, and recent evidence proposes a tripartite division of labour among these regions: the PPA supports a landmark identity code representing a specific place, RSC retrieves spatial and conceptual information about that particular place, and OPA represents perceptual details such as building materials and windows (Marchette, Vass, Ryan, and Epstein, 2015). These results emphasize the importance of examining these three scene-selective regions in conjunction to explore their contributions to visual perception. In the present dissertation, the evidence clearly demonstrates dissociations between these three scene-selective regions, but additional research is needed to clarify the nature of these dissociations and how visual feature information is processed in relation to the division of labour across scene-selective regions.

The influence of object information in shaping neural representations of scenes

Although LOC has been defined here as an object-selective region, this should not devalue the contributions it makes to scene perception. Indeed, LOC is involved in the processing of scene-like relations and interactions between objects (Kim & Biederman, 2010), as well as changes of the relative positions of objects (Hayworth, Lescroart, & Biederman, 2011),
suggesting this region plays an important role in scene processing. An intriguing finding from Chapter 2 concerns the modulation of activity in the PPA by the presence of an object within a scene. Specifically, attention to object texture resulted in greater activation in the PPA compared with attention to object shape when scenes and objects were presented separately (Experiment 1), but this difference was weakened and was no longer significant when scenes and objects were presented together (Experiment 2). Previous behavioural research has demonstrated an interaction between the processing of shared visual features (i.e., geometry and surface properties) across object and scene perception (Lowe et al., 2015), and neuroimaging investigations have further highlighted a role for object-selective cortex in modulating activity within the PPA (Mullin and Steeves, 2011; Rafique et al., 2015). These findings emphasize that feature perception in PPA is not static, but dynamic and dependent on the contextual relationship between the surrounding environment and the objects within it. The exact nature of the object-scene interaction observed in this study is unclear. Moreover, the general pattern of univariate results across Experiments 1 and 2 was qualitatively similar. As such, I hesitate to make strong conclusions regarding the nature of object-scene interactions from the present results. Factors such as ease of figure-ground segmentation, object size, location of the object within the scene, and congruent versus incongruent relationships between scene and object geometry or surface properties may all play a role in explaining object-scene interactions specifically and these results more generally.

A large body of research has found that the search for and recognition of objects is facilitated by knowledge about which objects and scenes tend to co-occur (Biederman, Mezzanotte, and Rabinowitz, 1982; Boyce & Pollatsek, 1992; Davenport, 2007; Davenport & Potter, 2004; De Graef, Christiaens, and d’Ydewalle, 1990; Henderson, Weeks, and Hollingworth, 1999; Joubert, Fize, Rousselet, and Fabre-Thorpe, 2008; Joubert, Rousselet, Fize, & Fabre-Thorpe, 2007; Palmer, 1975), and object information may also influence scene classification through a consistent-object advantage (Davenport, 2007; Davenport and Potter, 2004; Joubert et al., 2007; Mack and Palmeri, 2010). Such findings suggest a dual-system, interactive account between object and scene processing, and shared visual features across object and scene geometry and surface properties influences this interaction (Lowe et al., 2015). Here, evidence demonstrates overall greater sensitivity to changes in geometry over surface information in the object-selective LO, in both objects and scenes (Chapter 2) and in scenes.
without prominent objects (Chapter 3). Together, these findings suggest LO may play a role in processing both local and global geometric information in object and scene perception. At the very least, the results of Chapter 2 represent a first step in establishing the finding that placing an object within a scene modulates object-related, but not scene-related activity in the PPA. The degree to which this finding is explained by the various factors described above, and whether scene-related activity can be modulated by the presence of an object, are interesting empirical questions that deserve further study.

*Interactions of geometry and surface properties*

Visual features within our local visual environment can interact to support a common goal. For instance, surface properties (on both objects and scenes) may contribute to the perception of spatial information through cues provided from depth and contours (Torralba and Oliva, 2003), as well as edge information (Renninger and Malik, 2004), while geometric features may provide boundary information utilized in texture segmentation (Mohan and Nevatia, 1992). Support for the idea that geometry and surface properties are often used to accomplish a common goal comes from the results of Chapter 2, where the processing of scene shape and scene texture was found to be non-additive (Experiment 3). This suggests that these features are not represented independently in the PPA. For active navigation, surface properties provide affordance-related information pertaining to how an observer should move through an environment, and boundaries provided from geometric cues are similarly critical for navigation (e.g., obstacle avoidance). Indeed, how could an individual efficiently navigate across a landscape without knowledge of the material differences between snow and ice, or sand and grass, and the most efficient path in space to move from one place (i.e., the origin) to another (i.e., the goal)? The results presented here suggest that neural representations of these features may not be spatially independent within the PPA, but interestingly, there is evidence to suggest that these features may be temporally distinct (Chapter 4).

*The time-course of scene categorization*

In Chapter 4, findings indicate that scene categories can be decoded from neural activity within a fraction of a second within the human visual system. Notably, temporal signatures for
both global scene properties (scene content, spatial boundary), and basic-level categories (beach, forest, mountain, city, highway, office) emerged from neural activity within the first 100ms of perception, for both line drawings and colour photographs of scenes. These findings provide evidence that fundamental information required for perception of the visual environment is available even earlier than previously believed (e.g., Thorpe et al., 1996), and further reveal evidence that visual properties within a scene influence this time course. Critically, these properties emerged from neural activity during an orthogonal memorization task, supporting previous suggestions that these properties form a concrete and fundamental basis to our understanding of the visual world and may be automatically extracted (Oliva and Torralba, 2006).

Real-world scenes contain a myriad of rich information, such as structure and surface properties. These visual features interact across object and scene perception (Lowe et al., 2015), and directing attention to these features separately has been shown to influence the processing of global scene properties (Chapter 3). Within the spatial domain, disentangling the processing of these features has been a challenge, as they may be processed by similar brain regions (Walther et al., 2011; Park and Park, 2017; Chapters 2-3). Turning to the temporal domain is therefore critical to understand how the human visual system processes complex environments in a single glance. The results presented in Chapter 4 show that the processing of line drawings and colour photographs diverges along multiple dimensions across the time course of scene categorization, providing evidence that the relationship between the processing of scene features and scene perception may be well described within the temporal domain. Discrimination of global scene properties in line drawings and colour photographs temporally diverged over parietal and frontal regions, suggesting that the visual information available to observers may influence the time course of scene perception. Specifically, the time course of categorization in early occipital cortex was similar between line drawings and colour photographs, which may highlight the role of structure, which is preserved in both image formats (whereas surface properties are only preserved in colour photographs), as sufficient for the perception of both global scene properties and basic category levels. In contrast, differences emerged across line drawings and colour photographs over parietal cortex, revealing that surface information such as colour and texture may be utilized differently in later-stage processing.
Notably, the structure preserved in line drawings, and the subsequent ability of the visual system to process both global scene properties and basic category levels in line drawings suggests that structure may be a fundamental property of scene perception. This suggestion is supported by recent evidence showing that contour junctions underlie neural representations of scene categories (Choo and Walther, 2016), and evidence linking deficits in scene perception from topographical disorientation to a primary reliance on structural properties (Robin et al., 2017). Together with the present findings, this highlights an integral role for structure in perception, consistent with the proposal that edge-detection is primary to surface perception (Biederman, 1987; Biederman and Ju, 1988). Surface characteristics such as texture may facilitate visual search by delineating edge information necessary for the extraction of meaning, and may interact with edge information to form a stable representation of the visual world. Texture and colour information may also be instrumental in mediating scene recognition (Castelhano et al., 2008; Goffaux et al., 2005; Oliva and Schyns, 2000; Steeves et al., 2004; Renninger and Malik, 2004), and these findings highlight this relationship through dissociations between line drawings and colour photographs across both time and space. Scene categorization is influenced by diagnostically useful information (Malcolm, Nuthmann, and Schyns, 2014), and these visual features may therefore influence scene categorization in distinct ways. Here, the findings suggest that the structural information preserved in line drawings is sufficient for early scene categorization, yet these results further support a role for surface information in later-stage processing. Since these results did not show significant correlations across colour photographs and line drawings within the first 250ms of activity, they also suggest that low level image statistics may form a reliable and fundamental basis for scene categorization in the human brain (and may interact with higher-order, later-stage feedback processes, as discussed previously). These findings are consistent with previous evidence highlighting statistical regularity for real-world scene exemplars from within the same categories (Oliva and Torralba, 2001), and would account for the strong differentiation of scene categories over the occipital pole in the present results.

The findings of Chapter 4 also provide insight into a heated debate surrounding the temporal relationship between global scene properties, which represent the meaning of a scene, and basic-level categories, which represent the most common category descriptors. This debate concerns the hierarchical nature of scene processing, and which distinction (basic or global)
emerges first in the visual processing stream. Behavioural evidence has suggested that global properties of a scene (e.g., scene content) may emerge prior to even basic-level distinctions (Greene and Oliva, 2009; Loschky and Larson, 2010; Kadar and Ben-Shahar, 2012; Sun et al., 2016). In contrast, some evidence suggests that basic-level distinctions emerge prior to global categorizations (Rosch et al., 1976; Tversky and Hemenway, 1983). The results presented here suggest these properties may be discriminated from neural activity within a similar time window within the visual processing stream: Global and basic category level distinctions could be differentiated from activity within the first 100ms of scene processing over early visual areas for both colour photographs and line drawings. Neural markers for these distinctions in the present study therefore support behavioural evidence suggesting these properties may require the same amount of information in recognition (Fei-Fei et al., 2007). It is also important to note that these results provide evidence that the information contained within early neural signatures can be used to discriminate between scene categories, yet behavioural performance may differ from this temporal scale. For instance, while the information used to distinguish between scene categories may be present as early as 100ms, human performance and recognition may be linked to task context and observer goals (Chapter 2). One question for future research should therefore examine the extent to which task context influences how these different properties emerge during behavioural performance.

**Placing scenes within the broader context**

We are exposed to a great array of everchanging scenes, from the vast and sweeping landscapes of a desert, to the soaring monoliths of modern cities, yet we can identify our surroundings in only an instant. Despite the apparent ease with which the human visual system processes complex visual imagery, identifying the mechanisms underlying this process is not an easy feat. Throughout this dissertation, I have alluded to an ongoing debate surrounding the structure of functional representations within the PPA, and scene-selective cortex more generally. The spatial layout hypothesis, which posits the PPA represents the geometric structure of a scene as defined by its background elements (Epstein and Kanwisher, 1998), has remained a prominent theory for the structure of representation within this region. In the strongest version of this hypothesis, the PPA represents only geometric information, and eschews information about
texture, colour, and material properties (Epstein, 2014). Yet, in addition to the encoding of geometry, results provided in the present dissertation support a role for the PPA in the encoding of non-geometric visual properties of a scene, such as surface information. Previous investigations have further supported a role for the PPA in the encoding of low-level visual properties, such as spatial frequency content (Rajimehr et al., 2011; Watson et al., 2016), and other work suggests the PPA represents contextual associations of items which typically co-occur (Aminoff, Gronau, and Bar, 2007). This latter theory suggests that spatial and nonspatial contextual associations are processed by an anterior-posterior subdivision of the parahippocampal cortex, which houses the PPA.

In support of this subdivision of labour, object-sensitivity within the PPA was found to occur along an anterior-posterior gradient, with stronger responses to abstract object information, such as abstract sculptures, contained within posterior portions of this region (Baldassano, Beck, and Fei-Fei, 2013). Within the broader context, this anterior-posterior gradient may support a division of labour for nonspatial and spatial processing, respectively, and this process may jointly contribute to the formation of higher-level properties, such as scene identity and category. This theory would support a role for both geometry and surface properties in shaping scene identity and placing scenes within a broader context. For instance, while geometry and surface information could not be discriminated from one another across multiple experiments and stimulus sets in the present set of studies (Chapter 2 & 3), there is the possibility that spatial resolution is a limiting factor, and at a finer spatial scale, neural representation within the PPA could differentiate these features. An anterior-posterior subdivision of labour within the PPA could potentially tease apart these visual features. Moreover, a division of labour may not only occur across space, but also across time. In this dissertation, I have presented evidence for divergent processing of line drawings and colour photographs of scenes within the temporal domain, suggesting visual properties of a scene are not only separated in space, but also in time. Line drawings, which contain prominent structural properties with surface information such as colour and texture removed, are sufficient for scene categorization at multiple levels of processing (Chapter 4). These findings support the notion that edge-based information may receive priority processing for scene categorization (Walther et al., 2011), and later-stage processing may integrate surface information with geometry. Together, the present set of results
suggests scene recognition is jointly supported by the processing of multiple visual properties, which integrate across space and time with higher-level representations of scene identity.

**Limitations & Future investigations**

Ongoing discoveries continue to build upon a solid foundation of research to further our understanding of the human visual system, but there is still much to be investigated to refine this knowledge. An important question for future work involves identifying the individual roles and subdivision of labour of multiple regions within ventral visual cortex responsible for scene recognition. Critically, future investigations should examine the spatio-temporal dynamics of the processing in these regions to explore both similarities and dissociations in how they represent scene information across space and time. Precisely how the interconnectivity of these regions shapes their representations of visual feature information is also essential, and thus functional connectivity analyses should be used to examine feedforward and feedback connections across visual cortex. These investigations may shed light on how the human visual system integrates low-level visual properties with abstract meaning and identity. To circumnavigate these questions, we will need to investigate not only which visual properties are processed within these regions, but also how these properties contribute to scene identity.

The contribution of additional regions to scene perception should also be examined, as the processing within these regions may be equally important as the processing within the functionally-defined regions examined in this dissertation. For instance, the ability to navigate through an environment requires a cognitive map, or a mental representational structure that encodes spatial locations within a navigable environment in an allocentric manner (O’Keefe and Nadel, 1998). Individual differences in hippocampal volume predicts the ability to use spatial knowledge to make inferences about relative position, suggesting the hippocampus is the anatomical substrate mediating the ability to form and use cognitive maps (Schinazi et al., 2013). In line with this work, patients with hippocampal damage have difficulty discriminating images of scenes sharing overlapping features (Lee et al., 2005). Together, these studies suggest an essential role for the hippocampus in scene recognition and navigation, and perhaps even a role in visual feature discrimination. An important future question which ties into the present work should therefore examine how geometry and surface information is represented within the
hippocampus, and how this information influences the encoding of scenes. The present series of investigations has advanced our understanding of the neural mechanisms underlying scene perception and recognition and the visual properties responsible for shaping these mechanisms, yet there is certainly still much to be uncovered.

**Conclusions**

To date, there has been considerable debate surrounding the structure of representations in scene-selective cortex. While support for spatial encoding has been widespread, neural evidence for how surface information is processed within the context of a scene has received little study, despite computational, psychophysical, and neuropsychological evidence supporting a meaningful role for this feature in scene discrimination. In this dissertation, I have presented evidence indicating that scene-selective cortex represents multiple visual features, such as geometry and surface properties, and these visual features are shaped by perceived scene category and task context. Furthermore, while the spatial representations of these features were found to be similar within the PPA, they may be dissociated within the temporal domain and unfold across time to form a holistic percept of the natural world. The findings presented in this dissertation thus demonstrate the importance of multiple properties of visual information in shaping underlying neural representations of our local visual environment. Scene perception occurs with incredible speed, and the work presented here supports an efficient neural network able to discern multiple pieces of information from the environment quickly and effectively to extract meaningful information within only a fraction of a second. These findings tie together multiple factors (individual visual features, high-level scene attributes and categories, and task context), to highlight the fact that scene perception and recognition, and visual processing more generally, serves a particular goal which unfolds across neural space and time, and multiple aspects of our environment may contribute to this astonishing process.
References


Buckingham, G., Cant, J. S., and Goodale, M. A. (2009). Living in a material world: how visual cues to material properties affect the way that we lift objects and perceive their weight. *Journal of Neurophysiology, 102*, 3111-3118.


Good exemplars of natural scene categories elicit clearer patterns than bad exemplars but not greater BOLD activity. *PloS one, 8*(3), e58594.


Appendices

Appendix Item 1. Additional Examples of Experimental Stimuli. An example of each type of geometry (room shape) and each type of surface property (wall texture) used in Experiment 3, Chapter 2. Each room was displayed with all surface properties for a total of 100 unique stimuli.
Appendix Item 2. ERPs (N = 16) with 95% confidence intervals plotted for scene content (natural; manmade) with significant temporal clusters (asterisks) highlighted, indicating significant differences between natural and manmade scenes, *p < 0.01.

Appendix Item 3. ERPs (N = 16) with 95% confidence intervals plotted for scene boundary (open; closed) with significant temporal clusters (asterisks) highlighted, indicating significant differences between open and closed scenes, *p < 0.01.