Encoding and Decoding of Cortical Feedback to Human Early Visual Cortex

Andrew T. Morgan, M.Sc.

Submitted in fulfilment of the requirements for the
Degree of Doctor of Philosophy

Institute for Neuroscience and Psychology
College of Science and Engineering
University of Glasgow

October 2017

Containing studies performed at:

Centre for Cognitive Neuroimaging, Institute of Neuroscience and Psychology, University of Glasgow, Glasgow G12 8QB.

Scannexus, Oxfordlaan 55, 6229 EV Maastricht, Netherlands

Abstract Summary

Only 5% of excitatory input to primary visual cortical (V1) neurons corresponds to feedforward input from the retina, and only 20% of responses by these neurons can be explained by retinal input (Carandini, 2005; Muckli and Petro, 2013). Neuronal responses are therefore highly influenced by non-feedforward interactions, allowing the brain to combine external input from the retina with context and knowledge. This is accomplished by integrating feedforward input with signals from neurons processing higher-level or associative information. The signals transmitted from higher cortical areas to V1 are known as cortical feedback.

The neuroscientific community is in agreement that cortical feedback is an important aspect of brain processing. However, the information transmitted by feedback and what factors give rise to contextual feedback remain largely unknown. Feedback connections provide V1 neurons with information about their far surround receptive fields (Angelucci and Bressloff, 2006), and stimulation in the surround provides contextual information about the scene to non-stimulated portions of V1 (Smith and Muckli, 2010; Muckli et al., 2015). Occluded V1 activity patterns recorded using fMRI have been used to decode different scenes, but again, little is known regarding the nature and content of the contextual information that feedback transmits.

This thesis aims to examine the information in contextual feedback to early visual cortex, with a particular focus on V1. To investigate this topic we used an occlusion paradigm derived from that of Smith and Muckli (2010) and Muckli et al. (2015). During normal vision, both feedforward and feedback signals are present. As such, a useful approach to study feedback is to isolate it from feedforward input. We occluded one quadrant of the visual field during stimulus presentation in order to remove meaningful feedforward input about scenes in a portion of retinotopic visual cortex. We used fMRI to assess brain activity in early visual cortex, allowing us to detect dendritic signaling associated with cortical feedback due to its sensitivity to cortical energy consumption (Logothetis, 2007, 2008; Petro et al., 2014).

In Chapter 2, we investigated potential high-level information in cortical feedback to V1 and V2. We presented subjects with an expanded version of the occlusion paradigm from Smith and Muckli (2010) and Muckli et al. (2015). We included twenty-four partially occluded scenes from six categories and spatial depths. These two high-level scene characteristics were chosen because they have previously been shown to modulate early visual cortical responses (Walther et al., 2009; Kravitz et al., 2011). We were therefore interested in

whether these characteristics also modulate feedback to V1 and V2. We found that response patterns in these subregions contain high-level category information, but we did not find that visual depth information generalized across exemplars. Additionally, we found that retinotopic responses in Occluded V1 and V2 differed from each other, suggesting that feedback to these two areas has different information content, and matching the known anatomical connections from mid- and higher-level visual areas in the ventral stream (Rockland et al., 1994; Rockland and Ojima, 2003).

In Chapter 3, we probed the information content of Occluded V1 and V2 responses at multiple levels of complexity using Representational Similarity Analysis (RSA) and encoding models. By analyzing data from Chapter 2 in these frameworks, we were able to compare both local (voxelwise) and distributed (multi-voxel) Occluded responses to three biologically-inspired computational models (the contrast energy-based Weibull model, the orientation-based Gist model, and the mid-level vision H-Max model), and the high-level scene characteristics explored in Chapter 2. Using RSA, we also compared scene representations from Occluded and Non-Occluded areas. We found that in Non-Occluded areas, V1 and V2 represent scenes similarly, while Occluded V1 and V2 do not. We also found that scene representations in Occluded V1 and V2 were correlated with high-level Category and H-Max models. Individual voxel encoding models showed that Occluded V1 voxels within 5° visual angle of fixation encode low-level information about the occluded scene, while voxels outside of 5° encode higher-level information. These results highlight a potential visual field bias in the type of information transmitted to V1 through feedback, with foveal voxels receiving more precise, low-level scene information, and peripheral voxels receiving more invariant or global scene features.

In Chapter 4, we examined the laminar profile of Occluded V1 using high-resolution (0.8mm^3) 7T fMRI. We again expanded our stimulus set, now with 192 Occluded scenes and 192 Non-Occluded scenes. This large stimulus set allowed us to map scene information onto voxel responses in greater detail, and the use of both Occluded and Non-Occluded scenes allowed us to compare voxel responses when receiving only feedback with responses when receiving feedforward, lateral and feedback information. We found that V1 responses exhibit predictive and high-level response properties in addition to feedforward orientation and spatial frequency properties typically associated with V1 responses. These predictive and high-level responses were primarily associated with superficial layers of cortex. We also found that voxel tuning toward feedforward and feedback signals was different between cortical layers of V1. Our findings suggest that feedback connections terminating in superficial layers provide V1 neurons with contextual and associative information not available via localized feedforward input.

The neuroscientific results presented in this thesis extend our knowledge about the information content of cortical feedback to early visual cortex. These results add support to the notion that V1 can be considered to speak two languages (Muckli and Petro, 2013). Not only does it play a role as an early stage of processing of sensory visual input, where it deals with

processing low-level features, but it also receives messages from diverse areas of cortex and these messages supplement local processing by providing contextual information.

Contents

Abstract Summary i

List of Figures vii

List of Tables ix

List of Abbreviations x

Acknowledgements xi

Author's Declaration xii

Cortical Feedback to Early Visual Cortex 1

1 Introduction 1

1.1 Information processing streams in the visual system 2
1.2 Receptive field structure in primary visual cortex 6
1.3 Measuring cortical feedback signals using fMRI 8
1.4 High-level scene information in cortical feedback 10
1.5 Layers in cortex 10
1.6 Retinotopic mapping of early visual cortex 12
1.7 Modeling techniques for fMRI data 13
1.7.1 Population receptive field mapping 17
1.8 Thesis Aims 18

2 Cortical feedback to V1 and V2 contains unique information about high-level scene structure. 21

2.1 Introduction 21
2.2 Results 22
2.2.1 Decoding high-level cortical feedback to early visual cortex 23
2.2.2 Visualizing Retintopic Information Patterns 26
2.3 Discussion 28
2.4 Methods 30

2.4.1 Participants . . . 30
2.4.2 Stimuli . . . 30
2.4.3 Experimental Design . . . 31
2.4.4 fMRI Acquisition . . . 31
2.4.5 fMRI Data Preprocessing . . . 31
2.4.6 Multivariate Pattern Analyses . . . 32
2.4.7 Retinotopic projections . . . 33

3 Relating feedback to V1 and V2 to multiple levels of scene information. 35
3.1 Introduction . . . 35
3.2 Results . . . 36
3.2.1 Relating Scene Representations in V1 and V2 . . . 37
3.2.2 Voxel-wise Information Encoding . . . 39
3.3 Discussion . . . 42
3.4 Methods . . . 44
3.4.1 fMRI Data . . . 44
3.4.2 Representational Similarity Analysis . . . 45
3.4.3 Encoding Models . . . 45
3.4.4 Weibull Model . . . 46
3.4.5 Gist Model . . . 47
3.4.6 H-MAX Model . . . 47
3.4.7 Model redundancy measurements . . . 47
3.4.8 Absolute Depth Model . . . 48
3.4.9 Retinotopic projections . . . 48

4 Cortical feedback to superficial layers of V1 contains predictive scene information. 49
4.1 Introduction . . . 49
4.2 Results . . . 50
4.2.1 Depth-dependent information encoding . . . 51
4.2.2 Individual voxel tuning to feedforward and feedback signals . . . 54
4.3 Discussion . . . 55
4.4 Methods . . . 57
4.4.1 Participants . . . 57
4.4.2 Stimuli . . . 57
4.4.3 Experimental Design . . . 58
4.4.4 fMRI Acquisition . . . 58
4.4.5 MRI Data Preprocessing . . . 58
4.4.6 Population Receptive Field Mapping . . . 59
4.4.7 Encoding models . . . 60

4.4.8 Weibull Model 60
4.4.9 Gist Model 61
4.4.10 Category model 61
4.4.11 Retinotopic projections 62
4.4.12 Voxel tuning 62

5 General Discussion **63**
5.1 Decomposing the explanatory aspects of models 64
5.2 Deep learning models for studying feedback to early visual cortex 65
5.3 Studying cortical layer function in humans 67

Appendices **68**

A Supplemental Data for Chapter 2 **69**

B Supplemental Data for Chapter 4 **78**

List of Figures

1.1 Types of cortical connections. . . . 3
1.2 Visual pathway of the human brain from eyes to primary visual cortex. . . . 4
1.3 Feedback pathways carrying top-down information. . . . 5
1.4 Receptive field structure. . . . 7
1.5 Occlusion of natural scenes. . . . 9
1.6 Layer-specific fMRI. . . . 12
1.7 Retinotopic Mapping. . . . 13
1.8 Encoding and decoding in fMRI. . . . 14
1.9 Representational Similarity Analysis. . . . 16

2.1 Experimental procedures. . . . 23
2.2 Classification and cross-classification performance of Occluded and Non-Occluded V1 and V2. . . . 24
2.3 Representational Similarity Analyses for Occluded and Non-Occluded V1 and V2. . . . 25
2.4 Projections of V1 and V2 Linear Discriminant classifier weights into visual space. . . . 27

3.1 Similarity of six cortical regions of interest and computational models measured with Representational Similarity Analysis. . . . 38
3.2 Voxelwise unique model information encoding. . . . 41
3.3 Projections of V1 and V2 unique model information encoding into visual space. . . . 42

4.1 Creation of voxel-specific feature timecourses. . . . 51
4.2 Voxelwise unique model information encoding by cortical depth. . . . 52
4.3 Depth-specific projections of V1 voxels with significant model information encoding into visual space. . . . 53
4.4 Depth-specific feedforward and feedback voxel tuning. . . . 54

A.1 Voxel counts and density. . . . 74
A.2 Projections of V1 and V2 voxel responses into visual space. . . . 75
A.3 Absolute depth ratings of scenes. . . . 76

A.4 Individual-subject SVM decoding results. 77

B.1 Voxelwise unique model information encoding by cortical depth (unnormalized values). 79

List of Tables

A.1 Quantitative results from SVM classification analyses. 70
A.2 Quantitative results from SVM cross-classification analyses (lower-visual field occluder). 71
A.3 Quantitative results from SVM cross-classification analyses (upper-visual field occluder). 72
A.4 Subjects with significant individual subject classifications (permutation). . . 72
A.5 Subjects with significant individual subject classifications (Wilcoxon signed-rank). 73
A.6 Subjects with significant individual subject cross-classification (Wilcoxon signed-rank). 73

List of Abbreviations

3T . . . . . 3 Tesla

7T . . . . . 7 Tesla

BAC . . . . . Backpropagation-activated Ca2+ spike firing

EPI . . . . . Echo-planar imaging

fMRI . . . . . Functional magnetic resonance imaging

FFA . . . . . Fusiform Face Area

GLM . . . . . General linear model

iPAT . . . . . Integrated parallel acquisition techniques

IT . . . . . Inferior Temporal cortex

LGN . . . . . Lateral Geniculate Nucleus

LOC . . . . . Lateral Occipital Complex

MRI . . . . . Magnetic resonance imaging

MPRAGE . . . Magnetization-prepared rapid gradient echo imaging

pRF . . . . . Population receptive field

PPA . . . . . Parahippocampal Place Area

ROI . . . . . Region of interest

RSA . . . . . Representational similarity analysis

SVM . . . . . Support vector machine

V1 . . . . . Primary visual cortex

V2 . . . . . Secondary visual cortex

Acknowledgements

Firstly, I would like to thank my supervisor, Lars Muckli. The opportunities that you have given me have been monumental. Besides being able to conduct and engage in amazing brain research, I count 8 countries in which parts of this thesis were written, and I will always remember pulling over on the way to Salzburg to hike the Dragon Wall in conference clothes. Thank you to Lucy Petro; our countless conversations have been so enjoyable and taught me so much. I am lucky to have you as a colleague and as a friend. I would also like to thank Edythe London for the chance to work in her lab, for without it I would not be doing the research I am today. I learned so much from you and cannot thank you enough.

I would like to thank Fiona McGruer for her warmth and ability to convince me to trade time at my desk for climbing or running up a hill with her. Those reminders to be active have kept me sane during the last few years. I would like to express my gratitude to my lab members over the years for many exciting conversations - Grace, Matt, Yulia, Angus, Johanna, Fraser, Michele and Luca - I have thoroughly enjoyed this time. Thank you to Faruk, Marion, Francisco and Ingo for being so welcoming to me during my time in Maastricht, and especially for all the Thembi burgers. I would like to thank Federico De Martino for his guidance and support in learning the ropes of 7T scanning, and for great conversations while doing so.

I would like to thank Amaia Benitez for her support and willingness to test whether countless passages of my writing make any sense. You give so much of yourself to others - I am very lucky to have met you and have you in my life.

Most importantly, I need to thank my Mom. Without her unwavering support, I know that none of what I have done would have been possible.

Author's Declaration

I declare that, except where explicit reference is made to the contribution of others, this thesis is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution.

_____

Andrew Morgan

Chapter 1

Introduction

The conceptualization of the brain as a computing machine coincides with the invention of modern computers during the 1940s. Prior to the current description of brain function in terms of information processing and algorithmic computations, other then-current technologies were used to describe human intelligence. For instance, during the 17th century, Thomas Hobbes described thought as being the result of tiny mechanical processes in the brain. During the 18th century, this was updated to electrical processes, and during the 19th century, Helmholtz compared brain processes to the technology used to send telegraphs (Zarkadakis, 2015). Each of these brain descriptions reflected the most advanced technologies of their time. Thinking of brain function as computer-like processes has led to "profound physiological implications" for the study of cortical processing and to substantial gains in our understanding of brain function (Kandel, 2009). Nevertheless, our brains are not computers, and computers are not capable of performing many of the tasks that brains are capable of. Some of these tasks are complicated ones, requiring abstract thought processes and creativity. Surprisingly though, some are tasks that we find relatively simple or intuitive, such as catching a ball or describing a social context despite limited information.

Within a decade of modern computer development, studies began the pioneering work of mapping details from organisms' environments onto neuronal responses in order to test whether sensory neurons are selective in what information they process, much like specialized computer sensors and hardware. Horace Barlow (1953a,b) showed that certain neurons in frogs' brains fired only in response to specific visual stimuli. Hubel and Wiesel (1959) famously described spatially-localized receptive fields and orientation tuning of early visual cortical neurons in cats. Barlow (1961a,b) further laid out a theory of sensory systems in the brain, largely statistical in nature, which assigned sensory neurons the function of filtering external input in order to reduce redundancy. These influential works helped to form the still-current textbook account of early visual cortex, particularly primary visual cortex (V1), as a bank of oriented line detectors, called Gabor filters (Lee, 1996; Olshausen and Field, 1997). V1 neuronal responses would therefore represent visual input as oriented lines, and downstream brain areas could access these representations for use in more complicated

image computations.

This line of thinking, however, is focused on feedforward cascades of information processing (Petro et al., 2014). It largely neglects that the majority of connections to V1 neurons are not feedforward connections from the retina relayed through the Lateral Geniculate Nucleus (LGN) of the Thalamus, but are from other cortical neurons via feedback and lateral connections (Budd, 1998; Kayser, 2004). In fact, only 5% of excitatory input to V1 neurons corresponds to feedforward thalamocortical retinal input, and only 20% of V1 neuronal response variance can be explained by retinal input (Carandini, 2005; Muckli and Petro, 2013). V1 responses must therefore be highly influenced by cortico-cortical interactions occurring via feedback and lateral connections. Cortico-cortical influence, i.e. influence on cortical processing from other neurons within cortex, allows the brain to combine feedforward external input from the retina with context and knowledge by integrating feedforward and feedback signals. This integration is remarkably adaptive and flexible because it occurs at the level of local microcircuits throughout sensory and associative cortex rather than at a more centralized brain location (Mumford, 1992; Spratling, 2008; Bastos et al., 2012).

The significance of top-down and lateral influences on V1 neurons can be appreciated by considering cases when these cells do not strictly act as filter-like line detectors. For instance, V1 neurons respond differently to the same sensory line feature depending on whether that line appears in a natural or non-natural image stimulus (David et al., 2004). V1 activity also contains information about occluded stimuli (Sugita, 1999; Lee and Nguyen, 2001; Smith and Muckli, 2010; Muckli et al., 2015), about remembered scenes (Naselaris et al., 2015), and about complex sound stimuli (Vetter et al., 2014). These examples suggest that while much has been learned by mapping the feedforward aspects of neuronal activity, a complete understanding of neuronal physiology and brain function will require uncovering what types of information are conveyed by feedback and lateral interactions.

1.1 Information processing streams in the visual system

The human cortex performs increasingly complex computations in a hierarchical fashion. Artificial neural networks have successfully used hierarchical architectures inspired by sensory brain systems to perform many complex tasks, including scene and object classification, as well as speech recognition (LeCun et al., 2015). Generally, feedforward connections, also known as bottom-up connections, are those which project from an area specializing in lower-level features to one specializing in higher-level features. Lateral connections are those which connect neurons within a cortical area, and feedback connections, or top-down connections, are those which connect a cortical area specializing in higher-level features to one specializing in lower-level ones (Figure 1.1). The feedforward and feedback connections that facilitate these computations in the human cortex are well-established in neuroscience (Felleman and Van Essen, 1991; Markov et al., 2013a, 2014), but it is worth introducing the

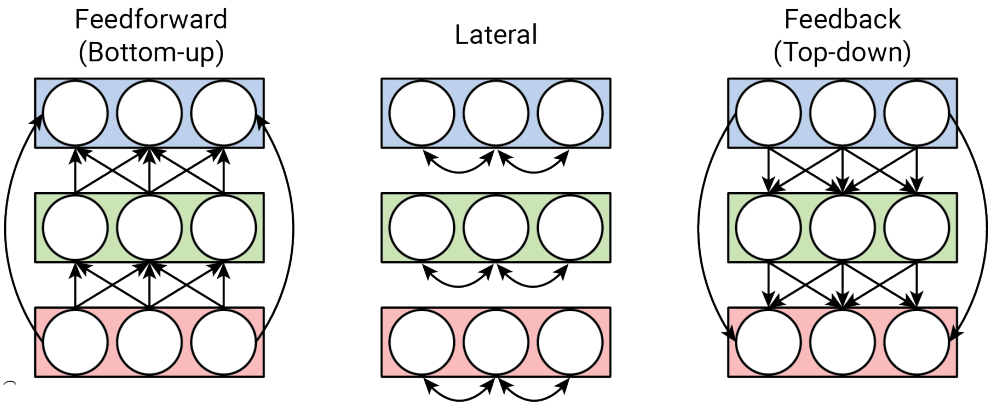

Figure 1.1: Types of cortical connections. Connection types are shown for three areas in an abstract cortical hierarchy, with red, green and blue areas representing low-, mid- and high-level areas, respectively (e.g. V1, V2, and V4). Circular nodes in these networks represent neurons or cortical microcircuits. Feedforward connections generally transfer information upward in the hierarchy, lateral connections allow nodes within an area to communicate, and feedback connections transfer information back down the hierarchy. Importantly, both feedforward and feedback connections can skip levels in the hierarchy. Non-feedforward connections refer to both lateral and feedback connections.

pathways through which information travels in the visual system from feedforward, lateral, and feedback directions to reach early visual cortical neurons.

Feedforward visual input is initially translated into neuronal signals by retinal photoreceptors - non-spiking neurons that transmit photometric information to multiple types of local spiking retinal ganglion cells (Hubel, 1988; Wässle, 2004). Individual photoreceptors are sensitive to incoming visual input from only a small, but contiguous portion of the visual field, and each receptor's area of sensitivity is known as its receptive field. Each eye contains approximately 125 million photoreceptors, but only 1 million ganglion cells. Ganglion cells therefore combine signals from many photoreceptors through direct and indirect pathways in topographically organized center-surround receptive fields. That is, they approximately decorrelate their outputs while conserving the local structure of their sensory input (Westheimer, 2004; Graham et al., 2006).

Through the optic nerve, the axons of retinal ganglion cells connect to various layers of the contralateral hemisphere of the LGN of the Thalamus, depending on the ganglion cell type they originated from (Figure 1.2). Continuing from this central-brain structure, axons from the LGN then fan out in a large band and project onto V1, which is situated at the Occipital pole and along the Calcarine Sulcus, a landmark sulcus on medial surface of the Occipital cortex. The topographic organization that was observed in the retina is conserved throughout this pathway (Hubel, 1988; Sereno et al., 1995; Brewer et al., 2005), thus, the spatial organization of V1 is termed *retinotopic*.

V1 is the primary target of visual input in cortex, receiving approximately 90% of signals sent from the retina (Douglas and Martin, 1991; Shao and Burkhalter, 1996). V1 then projects the majority of its output onto Secondary Visual Cortex (V2; Sincich and Horton, 2005). However, V1 also projects onto other higher visual areas, including V3-V5 (Figure 1.3). Projections from V1 span both the dorsal and ventral visual pathways, or the "where"

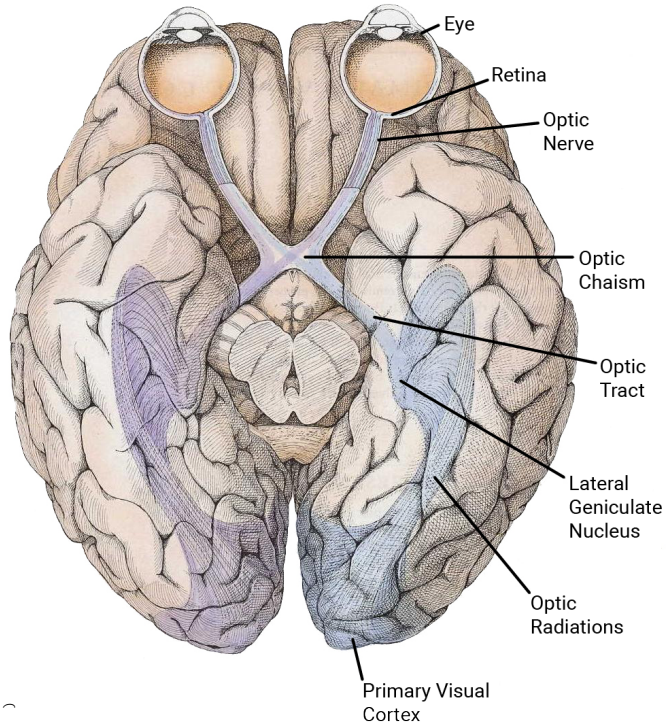

Figure 1.2: Visual pathway of the human brain from eyes to primary visual cortex. Information comes to the two purple-colored halves of the retinas (the right halves, because the brain is seen upside down) from the opposite half of the environment (the left visual field) and ends up in the right (purple) half of the brain. Reproduced with permission (Shepherd et al., 2013).

and "what" pathways according to the two-streams hypothesis (Goodale and Milner, 1992). Projections also exist to parietal and temporal cortices, although these projections are sparse (Borra and Rockland, 2011).

As one progresses anteriorly along the ventral surface of the Occipital cortex, cortical areas become increasingly specialized in processing high-level visual features (Grill-Spector and Malach, 2004). For instance, V2 preferentially responds to naturalistic patterns (Freeman et al., 2013), V4 responds to visual features of intermediate complexity (Tanaka, 1996), and Inferior Temporal (IT) cortex contains areas specialized in high-complexity features. These areas include the Fusiform Face Area (FFA), which preferentially responds to faces (Kanwisher, 2006), the Parahippocampal Place Area (PPA), which is involved in scene recognition and prefers scenes over objects (Epstein and Kanwisher, 1998), and the Lateral Occipital Complex (LOC), which is involved in object processing (Grill-Spector et al., 1999, 2001). The dorsal visual pathway is involved in processing motion, representing object locations, and informing movement of the eyes, hands and limbs. These complex processes recruit V2, the dorsomedial area, V5, and posterior parietal cortex (Goodale and Milner, 1992; Goodale, 2011).

As mentioned above, the sole direction of information passing is not upward in the visual hierarchy. In addition to feedforward input, neurons at any level of the visual hierarchy, from retinal ganglion cells all the way up to specialized neurons in IT cortex, also receive input from other neurons within the same area. Lateral connections in V1 allow neurons in retino-

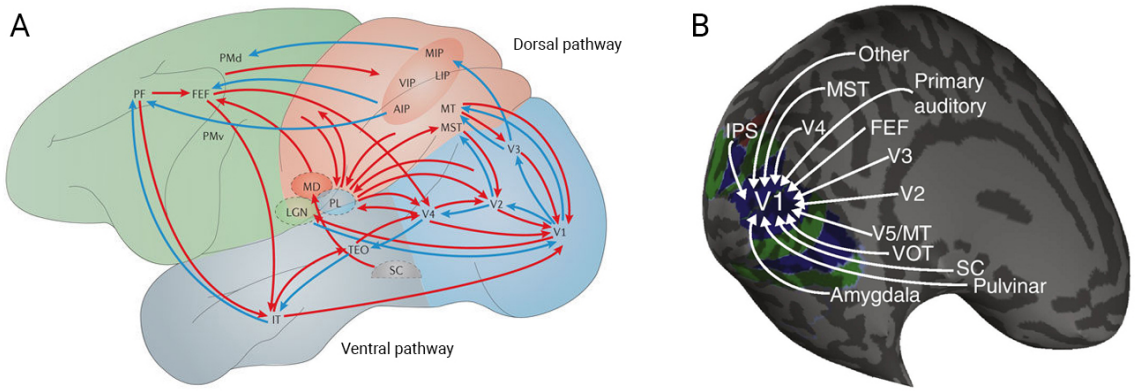

Figure 1.3: Feedback pathways carrying top-down information. Processing visual information involves feedforward connections across a hierarchy of cortical areas (represented by the blue arrows). The visual cortical pathways begin in the primary visual cortex (V1), which receives subcortical input from the lateral geniculate nucleus (LGN). The feedforward connections extend through a ventral pathway into the temporal lobe and through a dorsal pathway into the parietal cortex and prefrontal cortex (PF). Matching these feedforward connections are a series of reciprocal feedback connections (represented by the red arrows), which provide descending top-down influences that mediate re-entrant processing. Feedback is seen in direct cortico-cortical connections (those directed towards area V1), in projections from area V1 to the LGN and in interactions between cortical areas mediated by the pulvinar (PL). Reproduced with permission (Gilbert and Li, 2013; Muckli and Petro, 2013).

topically local neighborhoods to communicate, and can span up to 4.5mm along the cortical surface (or up to 5° visual angle; Angelucci et al., 2002). However, if information needs to travel laterally through multiple neurons (termed *polysynaptic* chains of lateral connections), then it is more efficient for information to travel upward and back down in the hierarchy via a shorter pathway consisting of feedforward and feedback connections (Angelucci and Bressloff, 2006).

Feedback connections allow for information passing downward in the visual hierarchy. Any given area of cortex receives feedback connections from about two-thirds of all brain areas. Such connections are composed of local reciprocal links, long range connections, thalamic and other subcortical connections (Figure 1.3; Larkum, 2013; Markov et al., 2013a). V1 receives feedback input from almost every area of the brain specialized in visual processing (Rockland et al., 1994; Rockland and Hoesen, 1994; Muckli and Petro, 2013; Petro et al., 2014). Additionally, V1 receives input from a number of associative and Parietal areas (Borra and Rockland, 2011), from subcortical structures (Markov et al., 2013a, 2014), and even from other sensory processing areas, such as auditory cortex (Rockland and Hoesen, 1994; Falchier et al., 2002; Rockland and Ojima, 2003; Wang et al., 2008). The abundance of non-feedforward connections to V1 neurons indicates that in addition to retinal input, V1 neurons are capable of integrating information from many different areas.

Some of the cortical feedback connections to V1 are likely transmitting aspects of sensory input that are not available via feedforward processing of visual input, for instance, information from other sensory modalities such as auditory or haptic information. However, internally generated activity, or neuronal activity not directly linked to sensory input, accounts for approximately 90% of cortical energy consumption (Raichle and Mintun, 2006;

Raichle, 2011). Such activity impacts cortical processing substantially, having been shown to be both perceptually and behaviorally relevant (Hesselmann et al., 2008; Takahashi et al., 2016). Further, these signals constitute our models and predictions about the world around us (Mumford, 1991, 1992; Muckli et al., 2013); models which can be described at multiple levels of abstraction - from statistical models of the likely relationships between adjacent visual patches (Lee, 1996; Gilbert and Li, 2013), to high-level visual models of scene categories (Oliva and Torralba, 2001, 2006; Walther et al., 2009), and even to conceptual beliefs about the self.

In light of the substantial impact non-feedforward signals have on V1 processing, it has been suggested that V1 models should incorporate cortical feedback signals in order to account for the wide response variance afforded to V1 neurons by their varied types of input (Angelucci and Bullier, 2003; Muckli and Petro, 2013). It is clear that this sizable task is only approachable by first systematically describing information conveyed by feedback. This is a significant challenge for neuroscience - predictions in the brain do not happen in isolation, but are instead constantly integrated with sensory input. Furthermore, cortical predictions likely range as widely in their complexity as the areas connecting to V1 (Mumford and Desolneux, 2010). It is therefore useful when dealing with visual input to encapsulate some of this complexity using high-level global properties such as scene category or scene depth. In this thesis, we approached the problem of describing feedback properties by isolating cortical feedback to human V1 using occlusion of natural scenes and recording activation patterns with functional magnetic resonance imaging (fMRI; Smith and Muckli, 2010; Muckli et al., 2015). This is possible because cortical neurons integrate feedforward, lateral and feedback signals at different scales. In the next section we outline the details of these differences.

1.2 Receptive field structure in primary visual cortex

Each V1 neuron responds to visual input from a specific region of the visual field, referred to as its classical, or central, receptive field. Within this preferred region, many neurons respond fervently to a particular visual feature, such as to a line of a particular orientation (Hubel and Wiesel, 1959). Importantly though, V1 neuronal firing rates can be influenced by visual features outside of their classical receptive fields, including near and far surround regions.

Figure 1.4 shows the structure of central, near-surround, and far-surround portions of a V1 neuronal receptive field. The LGN transmits highly localized visual information to the central portion of the V1 neuron's receptive field, approximately 1° of visual angle in diameter (Levitt and Lund, 2002). When presented in this center area, visual features which match the spatial orientation and frequency tuning properties of the neuron induce spiking. Surround receptive fields are referred to as the non-classical receptive field, as the presence of stimuli in the surround does not necessarily invoke spikes, but modulates responses to

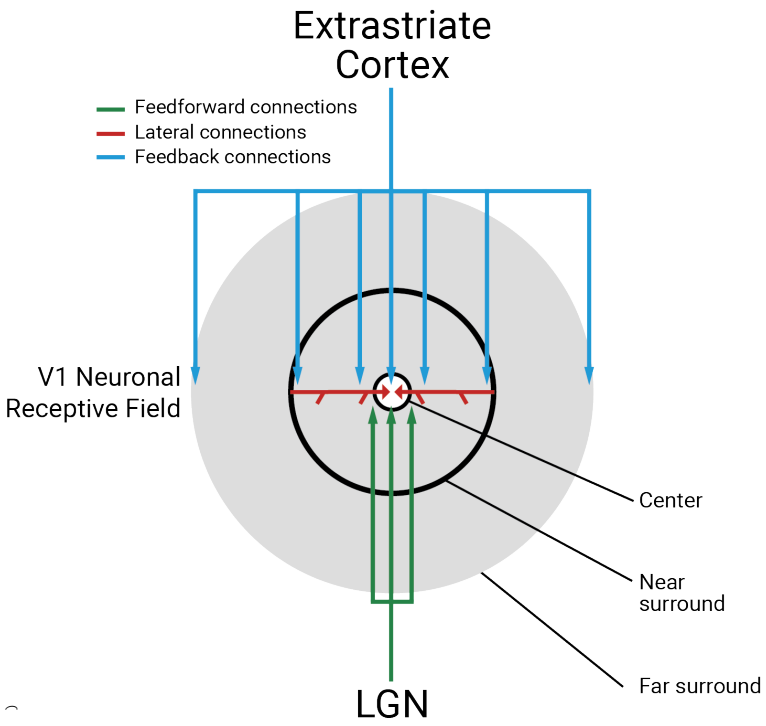

Figure 1.4: Receptive field structure. Schematic diagram showing the spatial scale of the different components of the receptive field center (white area) and surround (gray area) of V1 neurons and that of their underlying anatomical substrates. Adapted with permission from Angelucci and Bressloff (2006).

stimuli in the central, or classical, receptive field via dendritic mechanisms (Angelucci and Bullier, 2003; Angelucci and Bressloff, 2006; Gilbert and Li, 2013).

A neuron's surround receptive field is split into two different areas, the near and far surround, designating the types of connections influencing the neuron in that portion of its receptive field. The near surround receives input from other V1 neurons in the form of lateral connections, and neuronal responses are modulated by feedback connections in the far surround. Surround receptive fields are on average 4.6 times larger than their respective central counterparts, although this proportion depends on a neuron's eccentricity in the visual field (the visual distance from the center of the neuron's receptive field to the fovea; see Section 1.6). Between eccentricities of 2° and 8° of visual angle, V1 surround sizes are on average 5.1° of visual angle and can be as large as 13° in the periphery (Angelucci and Bressloff, 2006).

Receptive field sizes increase as one travels up the visual hierarchy (Kravitz et al., 2013), so feedback from higher areas conveys information about a larger region of the visual field compared to the classical receptive field. Additionally, it is important to note that the types of connections in each portion of the receptive field are not exclusive. In other words, the central receptive field receives feedforward, lateral and feedback connections, the near surround receives lateral and feedback connections, and the far surround is only influenced feedback connections. This organizational property of receptive fields means that isolating the signals coming from any type of connection to V1 neurons is difficult, traditionally requiring invasive methods such as pharmacological interventions, electrical stimulation, cool-

ing or optogenetics (Muckli and Petro, 2013). These methods commonly aim to moderate the non-feedforward signals reaching a particular neuron and then mapping modulations to feedforward responses.

Fully controlling the statistical properties of the large modulatory portions of a neuron's receptive field while stimulating or silencing remote neurons is a difficult task. An alternative strategy is to selectively homogenize feedforward signals across stimuli to the more easily-mapped central portion of the receptive field while allowing input to the surround portions (Williams et al., 2008; Smith and Muckli, 2010; Petro et al., 2013; Muckli et al., 2015). This approach effectively confines any differentiable aspects of the stimulus response to the non-feedforward signals arriving at the neuron. As previously mentioned however, non-feedforward signals often lack the direct spiking characteristics of feedforward input to V1 neurons (Larkum, 2013). Therefore, dendritic imaging techniques such as 2-photon microscopy are required to observe these signals directly (Denk et al., 1990; Helmchen and Denk, 2005), or methods sensitive to energy-consumption such as fMRI can be used to measure them indirectly.

1.3 Measuring cortical feedback signals using fMRI

Feedback to V1 influences wide-spread dendritic activity, but does not necessarily translate to spiking activity in cells (Olshausen and Field, 2005; Larkum, 2013). Functional MRI is particularly powerful for studying feedback because it densely samples large areas of the brain and is sensitive to energy-consumption (Logothetis, 2007, 2008). This means that its signal is influenced by spatially-specific dendritic activity rather than by the spiking output of neurons. Furthermore, fMRI is a non-invasive brain imaging technique, affording it access to feedback signals in humans. Since fMRI is capable of densely sampling visual areas of human cortex (Sereno, 1998; Brewer et al., 2005), it is possible to map the classical receptive field structure of thousands of individual fMRI voxels (these are the volumetric sampling units in fMRI - approximately 2-3mm^3 in 3T MRI studies, approximately 0.7-1.5mm^3 in 7T studies). The receptive field structures of voxels are termed *population receptive fields* (pRF; see Section 1.7.1; Dumoulin and Wandell, 2008) due to the summation of oxygenation-linked fMRI signal over an area containing a population of approximately 630 thousand neurons (3mm^3 voxel; Herculano-Houzel, 2009; Lent et al., 2011). pRFs are able to encapsulate the classical receptive fields of constituent neurons because a voxel contains neurons which are spatially localized in the visual field due to the retinotopic organization of visual cortex (Dumoulin and Wandell, 2008). Once voxel pRFs are calculated, it is possible to design studies which modulate voxel responses without stimulating the classical receptive fields of the contained neurons.

Section 1.2 stated that one strategy to study feedback is to selectively homogenize feedforward signals coming to the classical receptive field of neurons while allowing modulation

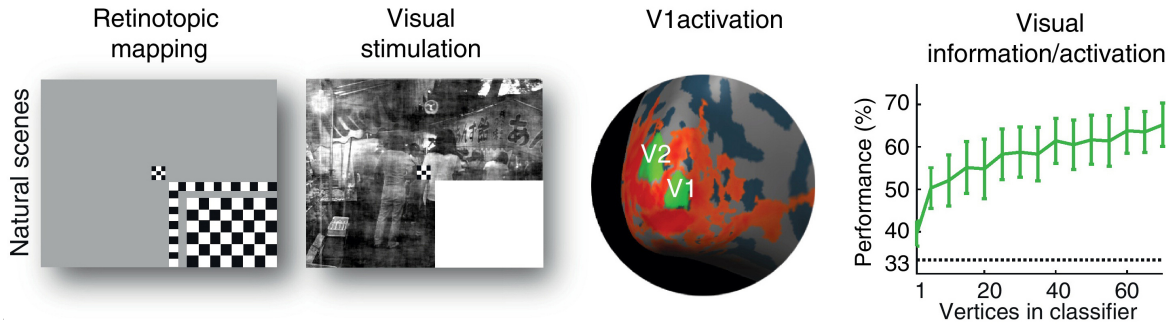

Figure 1.5: Occlusion of natural scenes. Experimental design of natural scene occlusion experiments in Smith and Muckli (2010) and Muckli et al. (2015). Cortical areas corresponding to the Occluded portion of scenes are mapped using contrast checkerboard mapping shown on the left, and only those voxels preferentially responding to the Occluded area compared to the surround of the Occluded area are included in analyses. Adapted with permission from Muckli and Petro (2013).

by non-feedforward input (Muckli and Petro, 2013; Petro et al., 2014). This approach has proven effective in experiments utilizing fMRI to study diverse forms of cortical feedback to V1 in humans, having uncovered modulation of V1 activation by auditory stimulation (Vetter et al., 2014), somatosensory stimulation (Liang et al., 2013), visual imagery (Vetter et al., 2014; Naselaris et al., 2015), attentional and task changes (Williams et al., 2008; Smith et al., 2010; Petro et al., 2013; Kay et al., 2015), and by contextual visual input to non-classical receptive fields (Smith and Muckli, 2010; Muckli et al., 2015; Kok et al., 2016). As mentioned in Section 1.1, this thesis examines cortical feedback based on occlusion of real-world scenes, utilizing the paradigm of Smith and Muckli (2010) and Muckli et al. (2015). The aforementioned paradigm presents subjects with images of real-world complex scenes which have their lower-right quadrant occluded by a homogeneous white occluder (Figure 1.5). The use of a white occluder rather than a background-matched grey occluder is thought to indicate to the viewer that visual information is hidden rather than missing, although experiments using this paradigm work in both cases. In these previous studies, the authors examined the response properties of cortical patches which retinotopically corresponded to the Occluded portion of the scenes (defined by independent visual mapping conditions) using multi-voxel pattern techniques (see Section 1.7). In this thesis, we have stringently modified criteria for classifying voxels as corresponding to the Occluded portion of scenes, thereby increasing confidence that examined voxels are not receiving meaningful feedforward input. Here, only voxels with pRFs lying completely within the Occluded portion of scenes are included in further analyses. Additionally, we have increased the number of scenes from 3 scenes from previous studies to 24 scenes in Chapters 2 and 3, and 192 scenes in Chapter 4. To probe the information characteristics of feedback signals coming to Occluded portions of early visual cortex, we chose scenes from different categories and scene depths.

1.4 High-level scene information in cortical feedback

David Mumford proposed a role for the reciprocal cortical pathways comprising of feed-forward and feedback information streams in which higher areas send predictions about the world to lower cortical areas (Mumford, 1991). At lower cortical areas, such as V1, top-down predictions would then be compared to sensory input arriving from the LGN. V1 receives feedback input from nearly every other visual area, including areas specializing in high-level visual processing (Rockland et al., 1994; Rockland and Hoesen, 1994; Muckli and Petro, 2013; Petro et al., 2014). Feedback signals sent to V1 are therefore conceivably related to predictions about multiple higher-level properties of visual input, thus providing abstract predictions that could bias processing at an early stage of visual processing (Lee and Mumford, 2003; Clark, 2013).

In this thesis, we compared fMRI voxel activity to biologically-inspired computational models spanning multiple levels of complexity. Models which would be considered feed-forward models of V1 included relatively low-level image statistics from a model of LGN processing known as the Weibull model (Scholte et al., 2009; Groen et al., 2013), and orientation and spatial frequency statistics via a V1-like model called the Gist model (Oliva and Torralba, 2001, 2006). More complex models, which can be thought of as feedback models include mid-level visual information (akin to V4 to anterior IT-level processing) from the H-Max model (Serre et al., 2007) and two high-level scene properties: scene category and depth. These properties were included because they have been shown to contribute to early visual cortical responses during normal vision (i.e. without occlusion; Walther et al., 2009; Kravitz et al., 2011).

We were interested in whether higher-level scene properties could be read out directly from feedback signals to Occluded V1. This would provide important detail about what types of information are present in cortical feedback. In computer science, feedback from higher to lower computational layers of artificial neural networks enhances their ability to represent high-level properties of visual input in conditions when input might be missing or degraded, such as during occlusion (Reichert and Serre, 2013; Spoerer et al., 2017). However, it is unclear what types of information are portrayed by the feedback channels in these networks. Such findings further highlight the importance of understanding the information content of feedback signals and how that information is integrated into computations in early processing stages like V1.

1.5 Layers in cortex

A critical aspect of the cortex is that it is not a two-dimensional homogeneous sheet, but is organized in layers. Different layers contain functionally distinct neurons, and importantly, feedforward and feedback connections terminate in different layers (Douglas and Martin, 1991; Felleman and Van Essen, 1991; Petro and Muckli, 2017). We have already established

the macroscopic routes by which information reaches early visual cortical neurons, so we now turn to a mesoscopic description of cortical layers. Artificial neural network models often refer to different processing stages as layers, but neuroanatomically, a given area of cortex (analogous to a single layer of a neural network) is commonly composed of 6 highly interconnected layers of tissue.

Feedforward connections typically terminate in middle layers of cortex, while feedback connections terminate in superficial and deep layers (Felleman and Van Essen, 1991; Markov et al., 2013a). Layer 5 pyramidal neurons are thought to be a primary integration site for combining forward and backward information streams in cortex. This is because their dendritic structure spans almost all layers and they compose a large proportion of cortical neurons (Larkum, 2013). Feedforward input to these cells is largely contained in middle layers. First, projections from the LGN terminate on layer 4 pyramidal neurons, which project onto layer 3 neurons, and finally input is fed to layer 5 pyramidal neurons through their basal dendrites (Douglas and Martin, 1991). Meanwhile, the apical tuft dendrites of these neurons are the primary targets of cortical feedback. Layer 5 pyramidal neuron cell bodies reside in the relatively deep layer 5 of cortex, hence their name, but their apical tufts reside in layer 1, the most superficial layer of cortex (Figure 1.6).

The distally located apical tuft dendrites of pyramidal neurons act as a separate integrative zone, somewhat independent from the integrative properties of basal dendrites of the same cells. These two distinct integrative compartments of pyramidal neurons appear to be involved in associative processing by allowing internal brain processes to provide contextual information to apical tuft dendrites that aids in processing of feedforward input to basal dendrites (Larkum, 2013; Häusser and Mel, 2003). Importantly, activation and deactivation of apical dendrites has been shown to directly influence perception in mice (Takahashi et al., 2016). The purpose of such anatomical segregation of functionally distinct connections is currently an open question; however, this cortical organization serendipitously allows experimentalists to probe feedforward and feedback signals.

Ultra high-field fMRI (using magnet strengths of 7 Tesla or higher) allows for non-invasive, functional imaging at multiple cortical depths in humans (Ugurbil, 2014). As mentioned in Section 1.3, many signaling processes in cortex are dendritic, which are linked to energy consumption, but not necessarily to neuronal spiking. fMRI is therefore a useful tool to measure these signals at a systems neuroscience level, and high-resolution 7T is a useful tool to disentangle particular laminar contributions. When studying feedback processes, this ability to differentiate signals from different layers can provide important information regarding the types of signals being imaged. For instance, 7T fMRI has shown that superficial and deep layers of cortex have larger receptive fields than middle layers (Fracasso et al., 2016) and that cortical feedback to superficial layers carries contextual information to localized areas of human visual cortex (Muckli et al., 2015).

In this thesis, we utilized both standard-resolution 3T (Chapters 2 and 3) and high-resolution 7T fMRI (Chapter 4) to study feedback properties in early visual cortex. The

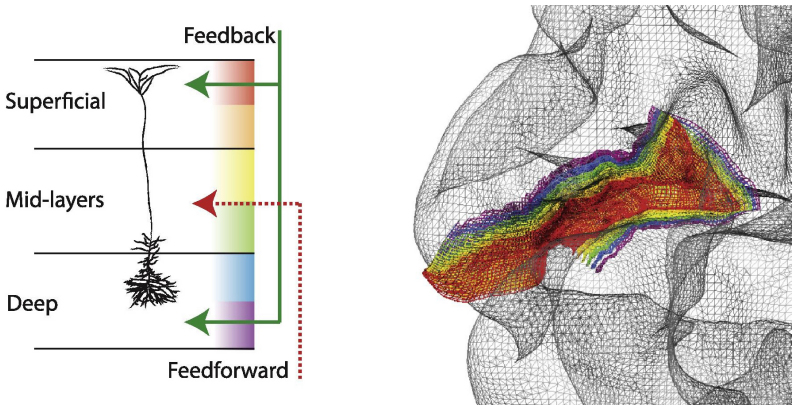

Figure 1.6: Layer-specific fMRI. Feedforward and feedback pathways are found in distinct layers of cortex, with feedback terminating largely in superficial and deep layers (green arrows) and feedforward in mid-layers (red dashed arrow). An example layer 5 pyramidal cell is shown, as this is a prominent target cell type of cortical feedback. Feedback arrives to the apical dendrites of pyramidal neurons in layer 1 (and to interneurons), whereas feedforward input arrives to the somatic region. Pyramidal neurons thus have two integration sites; one at the top of the apical trunk and one at the soma (see Larkum, 2013). Vertical color bars depict equidistant cortical depth sampling levels as has been studied with high-resolution fMRI of early visual cortex, such as that in the cortical model on the right. Adapted from Petro and Muckli (2017) under the Creative Commons Attribution License (CC BY).

recording and analysis of high-resolution 7T fMRI data is intricate; distortion and inhomogeniety issues in the signal and highly-involved segmentation processes currently make it difficult to conduct very large studies of mesoscopic processes. It was therefore useful for us to investigate how consistent feedback properties are using a large subject set in 3T fMRI (N = 23 across 2 experiments; 2 sessions each) before conducting longer and more individualized experiments using 7T fMRI (N = 3; 6 sessions each). The remainder of this introduction will briefly detail a number of important aspects of analyses used in this thesis.

1.6 Retinotopic mapping of early visual cortex

As described in Section 1.1, V1 straddles the Calcarine Sulcus of the Occipital cortex. Refining our description of the retinotopic organization of early visual areas, the deepest portion of this sulcus roughly corresponds to the horizontal mid-line of the visual field. Similar to the retina's inverted visual field representation, the lower half of the visual field image is represented by V1 in the dorsal Calcarine Sulcus, and the upper visual field is represented in the ventral portion of the Calcarine. Areas V2 and V3 wrap around V1 at the Occipital Pole (the most posterior portion of the brain). The lower visual field representation in dorsal V1 is reflected in dorsal V2, and once again in dorsal V3. The same occurs for ventral V1, V2 and V3 with the upper visual field. The fovea is represented at the Occipital Pole, and the far periphery of the visual field is represented by the most anterior portions of early visual cortical areas. More cortical area in V1 to V3 is dedicated to processing foveal portions of the visual field than is dedicated to peripheral portions. This organization is termed *cortical magnification* (Figure 1.7; Duncan and Boynton, 2003).

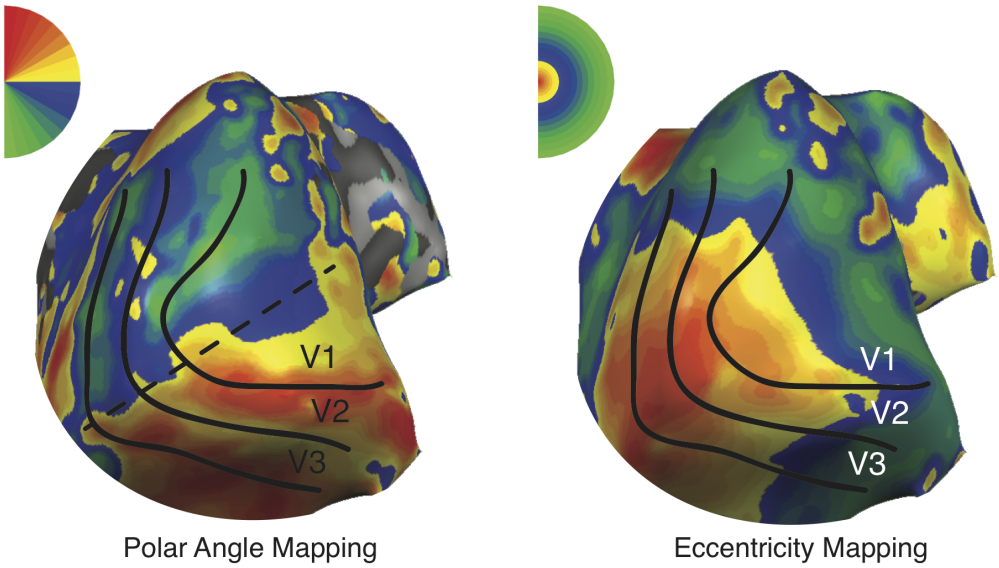

Figure 1.7: Retinotopic Mapping. Polar angle and eccentricity retinotopic mapping examples are shown on inflated Occipital Pole and Calcarine areas of the left hemisphere of an individual subject (from the data set collected in Chapter 2). Black lines signifiy V1 to V3 borders which are discerned from representations of the visual field. The dashed line on the polar angle map signifies the horizontal meridian of the visual field, and semi-circles signify visual field correspondence. Eccentricity mapping extended to 10° visual angle.

This pattern of representation can be mapped using retinotopically-specific visual stimuli in fMRI, and indeed was one of the first functional contrasts used (Turner et al., 1993). Two primary forms of retinotopic mapping stimuli are used to map early visual areas, corresponding to the angle and radius measures in a polar coordinate system. The first, called *polar angle mapping* consists of a radially-rotating wedge of high-contrast flashing checkerboard stimuli. The second, called *eccentricity mapping* consists of a checkerboard annulus which begins at a fixation point and expands outward into the periphery (Brewer et al., 2005). Each stimulus is typically presented as a "traveling wave" or "phase-encoded" stimulus, and is analyzed using Fourier or cross-correlation analysis (Engel et al., 1994; Sereno et al., 1995). By correlating the peak responses in voxel time courses to a particular stimulus phase, the polar angle and distance from fixation can be determined (Tootell et al., 1998; Sereno, 1998). Additionally, due to the representationally-reflective organization of retinotopic early visual cortex, this information can be used to deduce the visual area to which each voxel belongs.

1.7 Modeling techniques for fMRI data

A common goal of computational and cognitive neuroscience is to detect and describe information processing that occurs in the brain. To this end, we can think of two distinct and complimentary approaches to analyzing fMRI data. The first approach aims to predict stimuli based on brain activity, and models with this objective are termed *decoding* models. Decoding models are popularly referred to as brain-reading (Kessler and Muckli, 2011). The

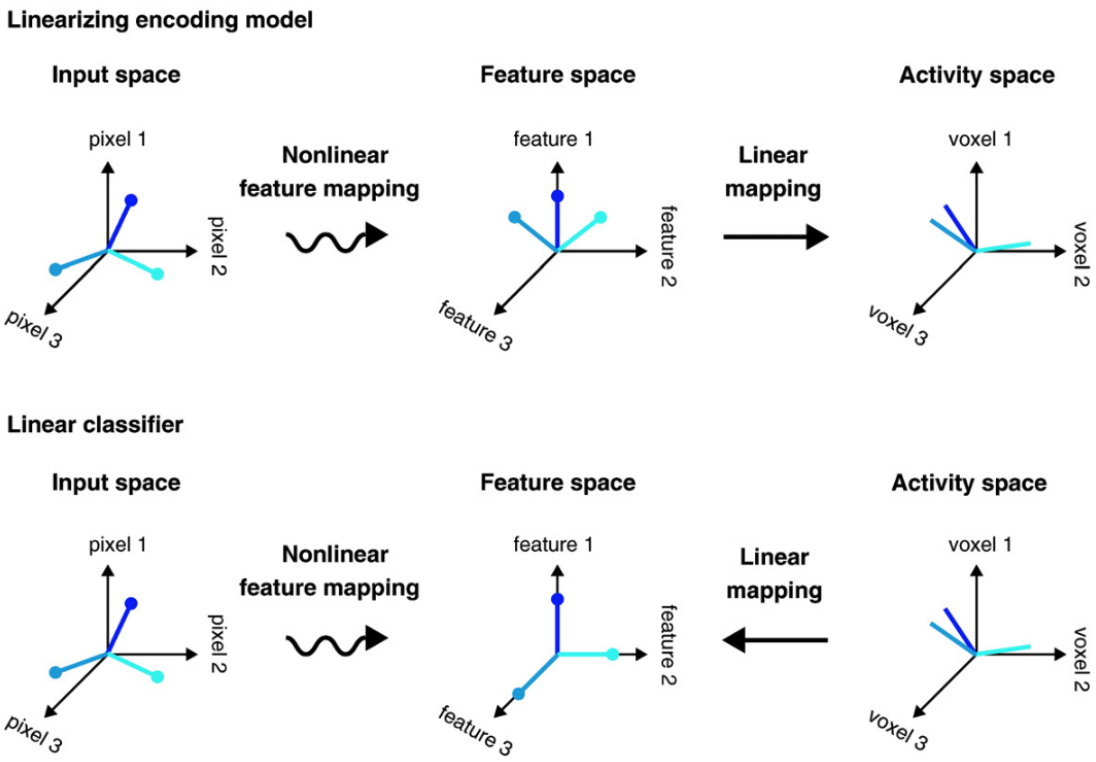

Figure 1.8: Encoding and decoding in fMRI. Encoding models: Pixel luminance is the input space of visual experiments, with each point representing a different image. Brain activity measured in each voxel is represented by an activity space, and between the input and activity spaces is a feature space. The mapping between the input space and the feature space is nonlinear and the mapping between the feature space and activity space is linear. **Decoding models:** The direction of mapping between activity and feature spaces is reversed compared to encoding models. Rather than features being mapped onto brain activity, brain activity is mapped onto the feature space. Adapted with permission from Naselaris et al. (2011).

second approach is the opposite; it aims to predict brain activity based on stimulus features, and models with this objective are termed *encoding* models (Figure 1.8; Naselaris et al., 2011). The studies contained in this thesis utilize a specific form of decoding model, a linear classifier. Additionally, encoding models were used to determine whether particular types of information are encoded by individual Occluded voxels (see Section 1.4).

Linear classifier models are a popular form of machine learning used in all branches of data science. When applied to fMRI data, classifiers are commonly referred to as Multi-Voxel Pattern Analyses (MVPA; Haxby et al., 2001; Kriegeskorte et al., 2008), and they aim to utilize the information contained in spatially-distributed activity patterns from fMRI. In MVPA, brain activity is analyzed at the level of patterns consisting of a number of voxels. This is important when studying feedback, as information in these signals could be distributed over large areas of cortex (Larkum, 2013; Muckli et al., 2015). Multivariate samples of brain activity are assigned labels indicative of the condition under which they were acquired. For example, the activity recorded while someone was viewing a picture of a cat or a dog would be labeled “cat” or “dog,” respectively (Haynes, 2015).

As is depicted in the right column of Figure 1.8, it is helpful to consider fMRI activity patterns in a multi-voxel activity space where the axes are individual voxel responses. In

this way, fMRI patterns can be thought of as vectors corresponding to the responses of all voxels taken together when an individual participant views a particular image (or any other type of experimental condition). Within this framework, it is easy to see that the distribution of responses between image categories (e.g. cat images vs. dog images) might not be separated within any of the single dimensions (single voxel responses). However, cat and dog responses could more easily be separated when all dimensions are taken into account. This is the task of classification algorithms: to find a way to separate categorical distributions while at the same time avoiding overfitting (Bishop, 2006).

One popular algorithm used to achieve fMRI activity classification is the linear Support Vector Machine (SVM) algorithm, which has been previously successfully employed in many fMRI studies for detecting different types of information in early visual cortex (Walther et al., 2009; Smith and Muckli, 2010; Smith et al., 2010; Vetter et al., 2014; Muckli et al., 2015). SVM uses a weight at each voxel to linearly project responses onto a single decision axis, which is defined as the maximum-margin hyperplane (Duda et al., 2000). That is, the algorithm finds the axis which maximally separates the individual data points from two classes. Training and testing of SVM classification models is performed on separate, independent data sets. This cross-validation framework strives to avoid overfitting to training data. Results are then reported as the number of testing data points that fall on their predicted side of the decision line (percentage correct). In addition, it is possible to train and test models on separate stimuli if stimulus sets are large enough to allow it. This method, known as *cross-classification*, avoids fitting models to specific exemplars from a category of stimuli. It is employed in Chapter 2 of this thesis when examining scene category and depth information in V1 and V2.

Classification generally requires training data to adhere to discrete groupings, each with a unique label. Scene descriptions such as category are discrete and thus easily modeled using classification methods. When scene descriptions are continuous, such as scene depth or the output from many computer vision models, these descriptions must be quantized or binned into discrete groups before decoding can be performed. However, binning continuous descriptors loses information, so it can be advantageous to use methods that preserve the information of continuous descriptors.

One such method is Representational Similarity Analysis (RSA; Kriegeskorte et al., 2008; Nili et al., 2014). RSA tests how well models fit distributed activity patterns in a given brain area by testing whether the similarity between brain responses to different stimuli matches the similarity between model responses to those same stimuli. This is done by computing the similarity (or dissimilarity) between brain or model responses to all stimuli. The choice between using similarity or dissimilarity is arbitrary, and might be dictated by the choice of metric. A common metric for describing the similarity of different stimuli is correlation (Pearson's r), and for dissimilarity, linear distance measures are common (such as Fisher's linear discriminant; Walther et al., 2016). The collection of all possible comparisons fully characterizes the "representational geometry" of the brain area or model in those stim-

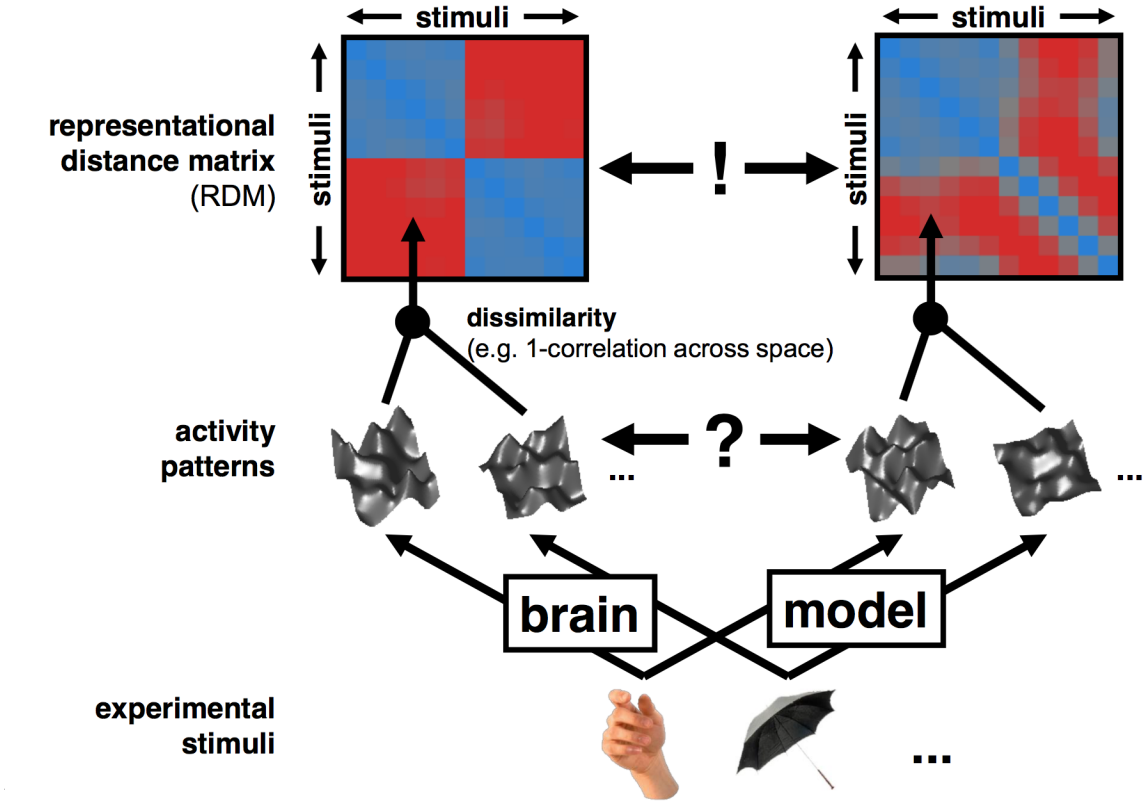

Figure 1.9: Representational Similarity Analysis. The activity pattern associated with each experimental condition is taken as a brain representation. By comparing activity patterns associated with each pair of conditions, we obtain a representational dissimilarity matrix (RDM), which serves to characterize the representation. Model comparisons can then be made by correlating brain and model RDMs (Kriegeskorte et al., 2008).

uli. These values are stored in a similarity or dissimilarity matrix, the indices of which can be compared to determine the similarity between brain and model representations (Figure 1.9).

Importantly, RSA is not limited to comparisons between brain recordings and computational models. This method is also capable of comparing brain recordings to behavioral responses, brain recordings from different modalities, different species, and different cortical areas. In Chapter 3, RSA is used to compare Occluded V1 and V2 to both computational models and to other brain areas (to Non-Occluded V1 and V2 and to each other).

As previously mentioned, a key difference between encoding and decoding models can be summarized by the direction of mapping between feature and activity spaces. Decoding aims to map voxel responses onto the feature space, either discretely or continuously in classification and RSA, respectively. Encoding models, however, map a feature space onto the activity space, and thus weights are assigned to stimulus features rather than to voxels. This mapping is most often performed independently for each voxel because regression methods commonly used to build these models are not well-suited for multidimensional data (for a demonstration, see Bishop, 2006). This means that when analyzing brain activity, decoding methods are convenient for describing information present in distributed patterns of activity over many voxels, and encoding methods are convenient for describing the stimulus features

that drive individual voxel responses[1].

Encoding models can be estimated from stimulus features of varying complexity. When mapping visual stimuli onto voxel activity, feature sets can range from simple pixel intensities (taken directly from presented images), to invertible changes of basis (Gabor pyramid decomposition or Fourier analysis), to nonlinear transformations such as the hidden layers of deep learning networks. The multitude of possible features for encoding makes these models incredibly powerful and flexible. This flexibility also means that encoding models need to be well-informed by theory. One example of the importance of theory-informed encoding modeling is pRF mapping, which we will discuss separately due to its importance in this thesis.

1.7.1 Population receptive field mapping

As discussed in Section 1.3, the term Population Receptive Field (pRF) refers generally to receptive field mapping, as in Hubel and Wiesel (1959), but derived from fMRI measurements. Adding the word 'population' to the name acknowledges that fMRI signals are influenced by the summed responses of many neurons (Victor et al., 1994; Dumoulin and Wandell, 2008). Briefly described, pRF mapping consists of mapping the response specificity of each voxel onto a stimulus space. In vision, this mapping is onto the visual field. A temporally- and spatially-varying stimulus is presented to the subject. Just as in other encoding regimes, the stimulus can have any level of complexity, and must span the parameters which the experimenter intends to assign to the voxel. For example, Kay et al. (2008) presented a large number of natural scenes to participants and assigned orientation, spatial frequency, position and size tuning characteristics to voxel pRFs. Natural scenes were used because when decomposed into Gabor filter responses, natural scenes span all of these characteristics in an efficient manner. Even still, many images are required to uncover all of these voxel response properties.

Many vision studies do not require orientation and frequency tuning information for individual voxels. Therefore, only spatial position and size parameters are mapped onto voxel responses. In this case, flashing checkerboards (as used for retinotopic mapping; see Section 1.6) are sufficient for pRF mapping. This is because retinotopic mapping stimuli elicit very strong retinotopically-specific visual responses even though they do not contain enough visual information to determine voxel orientation and frequency tuning. To efficiently model pRF maps with relatively small amounts of data (20 or fewer minutes of an fMRI experiment are typically dedicated to retinotopic mapping), pRFs are assumed to have an isotropic Gaussian structure (Dumoulin and Wandell, 2008). Under this assumption, it is possible to do a grid search of the parameter space to find the best fitting location and size model for each voxel.

[1]Methods are available for multi-voxel encoding and voxel-wise decoding, but are not common when analyzing fMRI data (Haynes, 2015).

In this thesis, visual pRF models aid our definitions of Occluded and Non-Occluded regions of interest (ROIs). We excluded any voxels that are not fully contained by both cortical surface and visual field ROIs. Based on the Gaussian structure of each voxel's model we can be confident that more than 97% of their feedforward responses are contained within 2σ of the pRF location. When building encoding models, pRF locations are also used to define relevant scene statistics for each voxel. By building feature sets using this prior knowledge about the response properties of voxels, we dramatically reduce the dimensionality of our models, and thus increase their statistical power.

1.8 Thesis Aims

The neuroscientific community is in agreement that cortical feedback is an important aspect of brain processing. However, the information transmitted by feedback and what factors give rise to contextual feedback remain largely unknown. Feedback connections provide V1 neurons with information about their far surround receptive fields (Angelucci and Bressloff, 2006), and stimulation in the surround provides contextual information about the scene to non-stimulated portions of V1 (Smith and Muckli, 2010; Muckli et al., 2015). Occluded V1 activity patterns recorded using fMRI have been used to decode different scenes, but again, little is known regarding the nature and content of the contextual information that feedback transmits.

The aim of this thesis is to examine potential levels of information in contextual feedback to early visual cortex, with a particular focus on V1. To investigate this topic we used an occlusion paradigm (see Section 1.3) previously used by Smith and Muckli (2010) and Muckli et al. (2015). During normal vision, both feedforward and feedback signals are present. As such, a useful approach to the study of feedback is to isolate it from the feedforward input. We therefore occlude one quadrant of the visual field during stimulus presentation in order to remove meaningful feedforward input about scenes in a portion of retinotopic visual cortex.

To assess brain activity in these regions we used fMRI, multi-voxel pattern analyses (MVPA) and encoding models. The use of fMRI allowed us to densely sample early visual cortex. Additionally, fMRI signals can detect dendritic signaling associated with cortical feedback due to its sensitivity to cortical energy consumption (Logothetis, 2007, 2008; Petro et al., 2014). MVPA allowed us to decode high-level features about the scenes surrounding occluded portions of the visual field from distributed brain activity patterns and to compare scene representations to biologically-inspired computational models of the occluded portions of scenes. The use of encoding models allowed us to predict individual voxel responses based on scene information ranging in complexity from low-level contrast energy or orientation filters to high-level global properties of scenes.

Experiments are described in **Chapters 2, 3,** and **4**. In **Chapter 2**, we investigated potential high-level information in cortical feedback to V1 and V2. We presented subjects

with an expanded version of the occlusion paradigm (Smith and Muckli, 2010; Muckli et al., 2015). This paradigm included twenty-four partially occluded scenes from 6 categories: Beaches, Buildings, Forests, Highways, Industries, and Mountains. Two exemplars from these categories were selected to be near the viewer, and two were farther away. These two high-level scene characteristics, Scene Category and Scene Depth, were chosen because they have previously been shown to modulate early visual cortical responses (Walther et al., 2009; Kravitz et al., 2011). We were interested in whether these characteristics also modulate feedback to V1 and V2. We found that response patterns in these subregions contain high-level category information, but we did not find that visual depth information generalized across exemplars. Additionally, retinotopic responses in Occluded V1 and V2 differed from each other, suggesting that feedback to these two areas has different information content. This finding matches the known anatomical connections from mid- and higher-level visual areas in the ventral stream (Rockland et al., 1994; Rockland and Ojima, 2003).

In **Chapter 3**, we probed the information content of Occluded V1 and V2 responses at multiple levels of complexity using Representational Similarity Analysis (RSA) and encoding models. By analyzing data from Chapter 2 in these frameworks, we were able to compare both local (voxelwise) and distributed (multi-voxel) Occluded responses to three biologically-inspired computational models. First, the Weibull model is a contrast energy and coherence model reminiscent of LGN processing (Scholte et al., 2009; Groen et al., 2012, 2013); second, the Gist model is a model of spatially-specific orientation and spatial frequency information, likened to V1 processing (Oliva and Torralba, 2001, 2006); and lastly, the H-Max model is a mid-level visual feature model with more spatial feature invariance, similar to V4 or anterior IT processing (Serre et al., 2007; Kriegeskorte et al., 2008). We also compared responses to the high-level scene characteristics explored in Chapter 2 (Category and Depth), and using RSA, we compared scene representations from Occluded and Non-Occluded areas. We found that in Non-Occluded areas, V1 and V2 represent scenes similarly, while Occluded V1 and V2 do not, confirming differences observed in Chapter 2. Further, RSA showed that scene representations in Occluded V1 and V2 were correlated with high-level Category and H-Max models, but were uncorrelated with the orientation-specific Gist model. Using encoding analyses, we showed that uniquely high-level Category and Depth information is statistically correlated with individual voxel responses. These analyses also uncovered relationships between voxel scene responses and the scene statistics hidden by our occluder. Furthermore, they show that across subjects, V1 voxels within 5° visual angle of fixation were uniquely related to low-level models (Weibull and Gist), while voxels outside of 5° were related to higher-level models. This highlights a potential visual field bias in the type of information transmitted to V1 through feedback, with foveal voxels receiving more precise, low-level scene information, and peripheral voxels receiving more invariant or global scene features.

In **Chapter 4**, we examined the laminar profile of Occluded V1 using high-resolution ($0.8mm^3$) 7T fMRI. We again expanded our stimulus set, now with 192 Occluded scenes

and 192 Non-Occluded scenes. This large stimulus set allowed us to map scene information onto voxel responses in greater detail, and the use of both Occluded and Non-Occluded scenes allowed us to compare voxel responses when receiving only feedback with responses when receiving feedforward, lateral and feedback. We found that V1 responses exhibit predictive and high-level response properties in addition to feedforward orientation and spatial frequency properties typically associated with V1 responses. These predictive and high-level responses were primarily associated with superficial layers of cortex. We also found that voxel tuning toward feedforward and feedback signals was different between cortical layers of V1. Our findings suggest that feedback connections terminating in superficial layers provide V1 neurons with contextual and associative information not available via localized feedforward input.

The findings from all three experimental chapters are discussed in **Chapter 5** in line with the current literature and I propose ideas for further research for the remaining unanswered questions.

Chapter 2

Cortical feedback to V1 and V2 contains unique information about high-level scene structure.

Early visual cortical neurons receive highly selective feedforward input, which is amplified or disamplified by contextual feedback and lateral connections. A significant challenge for systems neuroscience is to measure the feature space that drives these cortical feedback channels. We occluded visual scenes and measured non-feedforward stimulated subregions of V1 and V2 using fMRI and multi-voxel pattern analyses. We found that response patterns in these subregions contain high-level category information, but we did not find that visual depth information generalized across exemplars. Responses in non-feedforward stimulated V1 and V2 differed from each other, suggesting that feedback to these two areas has different information content. Our findings are consistent with descriptions of the visual system as a hierarchical inference network, with V1 acting as a high-resolution geometric buffer or blackboard to which feedback is projected.

2.1 Introduction

Cortical neurons receive two sources of input. The first, feedforward input, is sensitive to small portions of stimulus space as defined by classical receptive fields. The second consists of non-feedforward input from feedback and lateral interactions, which amplify and disamplify responses to the feedforward signals based on context (Gilbert and Li, 2013; Phillips et al., 2015). The abundance of feedback connections in cortical circuits (Budd, 1998; Muckli and Petro, 2013) and the diversity of responses to identical stimuli in different contexts (Petro et al., 2013) points to a challenge central to the understanding of neural computations. Improving our knowledge of feedback will provide insight to fundamental neuroscientific questions such as cognition, attention, perception, memory, action and mental disorders.

To study feedback, one must disentangle feedforward and non-feedforward sources of input. This involves independent stimulation or inactivation of feedback and feedforward pathways, which is achieved via single- or multiunit recordings paired with electrical stimulation, pharmacological intervention, cooling or optogenetics (Muckli and Petro, 2013; Petro et al., 2014). These methodologies provide detailed characterizations of neural activity, but are generally too invasive for studying the healthy human brain. Since the coordination of feedforward signals likely occurs through widely distributed feedback processes of contextual modulation (Phillips et al., 2015), non-invasive modalities with large spatial coverage such as functional magnetic resonance imaging (fMRI) are advantageous for studying feedback properties. Furthermore, fMRI is sensitive to energy-use rather than directly to spiking, and while feedback modulates feedforward responses, it does not necessarily generate spikes (Phillips et al., 2015).

A non-invasive strategy to measure feedback is to homogenize feedforward input using visual occlusion. Functional brain imaging studies have measured feedback-specific effects in retinotopic visual cortex (V1) by occluding feedforward input across experimental conditions and observing that feedback signals are heterogeneous (Muckli and Petro, 2013; Williams et al., 2008; Smith and Muckli, 2010; Muckli et al., 2005). Smith and Muckli (2010) blocked feedforward input to subsections of V1 using a uniform occluder and demonstrated that scenes can be decoded using pattern classification methods from responses in the occluded cortical area. Petro et al. (2013) showed that decoding of faces in regions of V1 that process facial features is dependent on task. Vetter et al. (2014) were able to decode different sounds from V1 activity in blindfolded subjects. These results provide evidence for feedback affecting V1 responses in a contextually relevant way, yet we can say little about the content of feedback signals. These signals may process specific feature predictions, and therefore produce response patterns similar to those observed when early visual cortex is stimulated by feedforward inputs (Naselaris et al., 2015). Conversely, feedback may provide early visual cortex with general templates for expected scene structure based on high-level characteristics. In the current study, we aimed to investigate the specificity of feedback signals to human primary and secondary visual cortices (V1 and V2) during scene processing.

2.2 Results

We blocked feedforward input to subsections of retinotopic visual cortex using a uniform visual occluder covering one quarter of the visual field (Smith and Muckli, 2010) while participants viewed 24 real-world scenes. To probe information characteristics of feedback, we identified two abstract features that modulate V1 responses: scene category and depth (Walther et al., 2009; Kravitz et al., 2011). Scenes were balanced across six categories (Beaches, Buildings, Forests, Highways, Industry and Mountains) and two spatial depths (Near and Far). We localized subsections of V1 and V2 responding either to the occluded

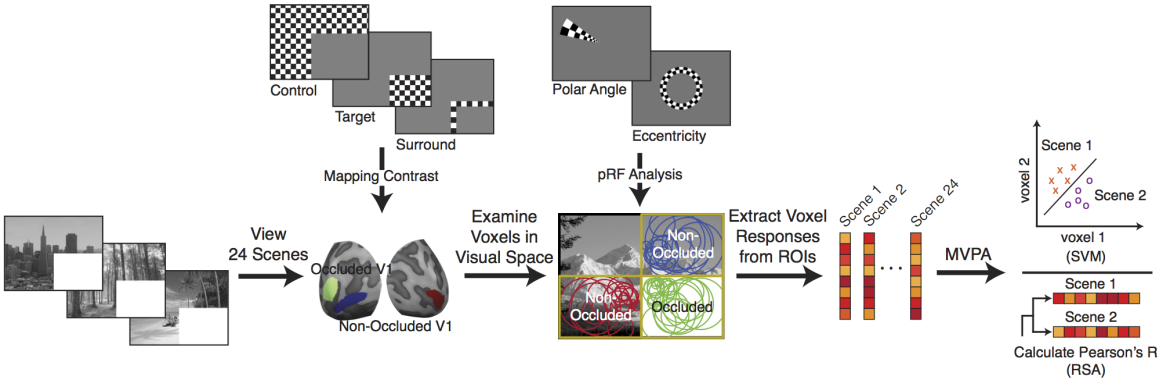

Figure 2.1: Experimental procedures. Participants viewed 24 scenes with lower right quadrants occluded. Scenes spanned 6 categories (Beaches, Buildings, Forests, Highways, Industry, and Mountains) and 2 depths (Near and Far). Occluded and Non-Occluded subsections of early visual cortex were localized using mapping contrasts. Retinotopic mapping data were used to separate V1 and V2 and for mapping population receptive fields (pRFs). Voxel pRFs not completely contained by the quadrant of interest (2σ from pRF center) were excluded from further analyses. Remaining voxels were included in multi-voxel pattern analyses.

portion of the visual field (lower-right image quadrant), or non-occluded visual field (upper-right and lower-left quadrants; Figure 2.1). This process yielded three regions of interest (ROIs) in each of V1 and V2, hereafter referred to as Occluded and Non-Occluded (either Upper-Right or Lower-Left), totaling six ROIs (3 positions in V1 and V2). We also mapped population receptive field (pRF) locations of individual voxels (Dumoulin and Wandell, 2008) to ensure that their response profiles were within regions of interest in the occluded visual field.

2.2.1 Decoding high-level cortical feedback to early visual cortex

Using single-trial linear Support Vector Machine (SVM) classification we could decode individual scene, category, and depth information in occluded and non-occluded regions of V1 and V2 (Figure 2.2, Table S1A; permutation tested against chance-level using 1000 samples, all p-values < 0.001). Occluded V1 and V2 were not significantly different when decoding category and depth. However, scene, category and depth decoding was higher in Non-Occluded V1 than Non-Occluded V2 in all analyses other than category decoding in the lower-left quadrant ($p < 0.05$). Our decoding results were reliable at the individual-subject level (Figure A.4); scene, category, and depth decoding were above chance-level in at least 14 of 18 subjects in nearly all regions tested (Table A.2). Only decoding of Depth in Occluded areas had lower individual-subject classifier performances, which were above chance-level in 10 and 7 (of 18) subjects in Occluded V1 and V2, respectively.

To further test whether non-stimulated V1 and V2 represent higher-level properties of scenes, we performed cross-classification analyses for category and depth information. We trained SVM models using responses to a subset of our scenes, leaving out a test-set for later cross-classification. For analysis of category, 18 (of 24) scenes were selected, leaving out one scene per category. For depth, we selected 22 (of 24) scenes, leaving out one scene per depth.

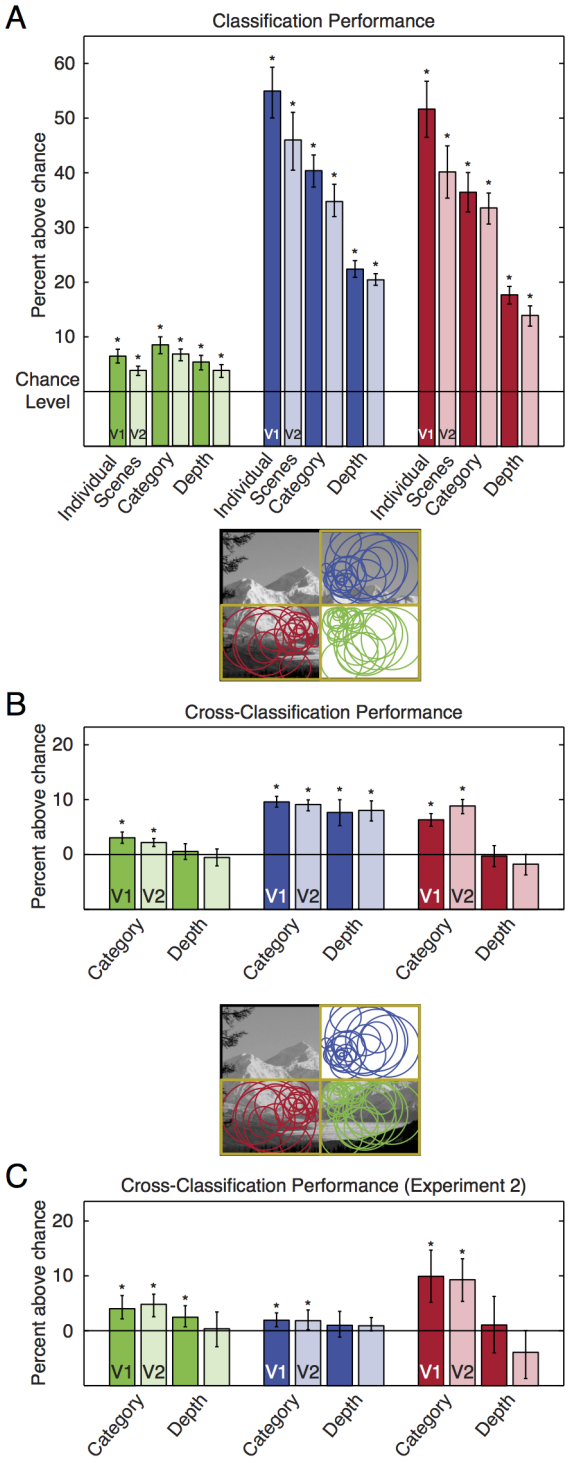

Figure 2.2: Classification and cross-classification performance of Occluded and Non-Occluded V1 and V2. (A) Average classifier performances (18 subjects) are shown for each visual area with 95% confidence intervals (calculated via 1000 bootstrap samples of individual subject performances). Occluded analyses are shown in green; Non-Occluded analyses are shown in blue. Asterisks indicate greater than chance-level decoding accuracy, $p < 0.05$. Chance level is 4.17% for individual scenes, 16.67% for categories, and 50% for depth. (B-C) Cross-classification performance for Experiments 1 and 2 (18 subjects in Experiment 1; 5 subjects in Experiment 2). Training occurred on 18 and 22 (of 24) randomly chosen scenes in category and depth analyses, respectively, and testing occurred on scenes not used for classifier training. This was repeated 100 times per subject, and 95% confidence intervals were calculated via 1000 bootstrap samples of individual subject performances. See also Table A.1, Table A.2, Figure A.1 and Figure A.4.

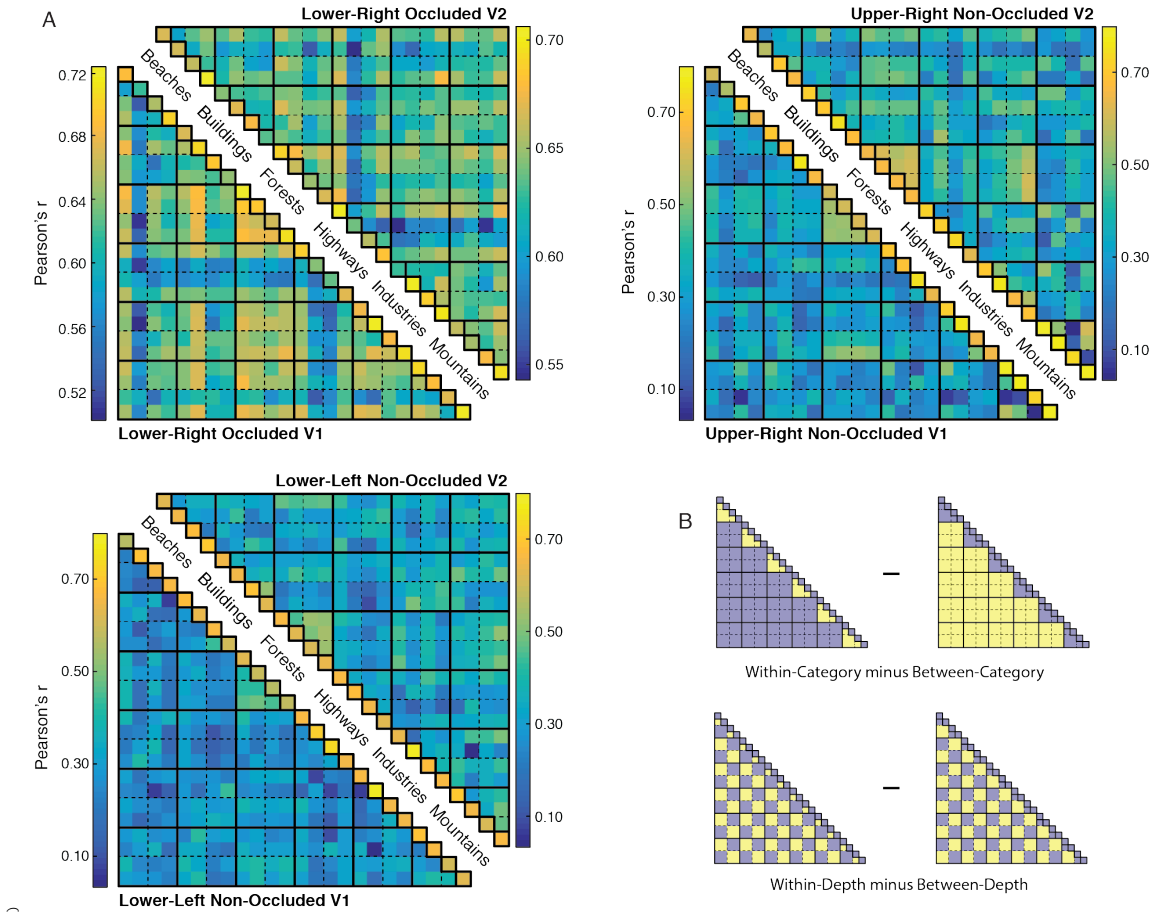

Figure 2.3: Representational Similarity Analyses for Occluded and Non-Occluded V1 and V2. (A) Similarity matrices comprise of correlations in the patterns of response between pairs of scenes, averaged amongst subjects. Diagonals (outlined in black) are correlations of scene responses in each half of the data; a measure of signal reliability. V1 is shown in the lower triangle of each matrix and V2 in the upper triangle. (B) The presence of high-level information in each RSA matrix is tested by comparing the average of all within- and between-Category (or -Depth) indices.

We tested the classifier on the remaining scenes in a cross-classification approach. Due to the large number of possible image permutations in these analyses, we randomly assigned scenes to training and testing sets 100 times in each subject. Cross-classification of category was successful in Occluded and Non-Occluded areas. Cross-classification of depth was only successful in the Non-Occluded Upper-Right quadrant, suggesting that depth information is not available in lower visual field responses regardless of whether feedforward information is available (Figure A.2, Table A.2).

Our cross-classification results show that responses in both Occluded and Non-Occluded areas of the lower visual field do not contain depth information. This visual field bias limits our interpretation about the presence of depth information in feedback to early visual cortex. We therefore conducted a second fMRI experiment in five subjects using the same scenes, but with the occluder moved to the upper-right quadrant of the visual field. In this experiment, we successfully cross-classified category in all areas, but once again failed to cross-classify depth information in the Occluded quadrant (Figure 2.2, Table A.3). Based on the cross-

classification results of our two experiments, we conclude that Occluded responses contain generalizable information about scene category, but not scene depth.

SVM classification and cross-classification allowed us to determine whether information about individual scenes, category, and depth was present in the responses to single presentations of images and whether this information was generalizable across scenes. Additionally, we aimed to detect Individual Scene, Category and Depth information in average response patterns by conducting Representational Similarity Analyses (RSA). Applying an iterative split-half cross-correlation procedure (Kravitz et al., 2011), we compiled matrices representing correlations between individual scene response pairs in each region (Figure 2.3). Diagonal points of similarity matrices indicate consistency in response patterns to individual scenes between the two halves of the data, and off-diagonal points are correlations between pairs of responses to different scenes. As such, higher correlations in the diagonal indicate the ability to distinguish individual scenes within a region of interest. Here, comparison of values along the diagonal against those off the diagonal showed significant differences in regions ($p<0.001$, two-sided t-test of Fisher-transformed correlation values). These results reaffirm that responses in Occluded subregions of V1 and V2 contain information about individual scenes. To test for category and depth information, we implemented a two-sided paired t-test comparing average within- to between-Category (and Depth) comparisons using Fisher-transformed correlation values (Figure 2.3). Corroborating our cross-classification results, we found that Category information is present in all areas, and that depth information is only present in the Upper-Right Non-Occluded areas (Table A.3). Overall, combined results from SVM and RSA analyses demonstrate that Occluded responses carry information about individual scenes and about scene category, but not about depth.

2.2.2 Visualizing Retintopic Information Patterns

Our Occluded V1 and V2 decoding results detect reliable differences in response patterns between conditions, but do not indicate what scene features elicited them. This consideration is particularly important in Occluded regions, where we would like to better understand the nature of feedback information. For example, we want to ensure that successful decoding is not simply due to spill-over or edge effects at the occlusion boundary. Additionally, we are also interested in whether information utilized by classifiers corresponds to continuations of objects or contours under the occluder. To visualize classification information over corresponding scene features, we projected voxel classifier weights (the relative contribution of each voxel to an LDA solution) into visual space (Figure 2.4). Positive weights indicate voxel responses associated with a classifier choosing the current scene and negative weights indicate responses associated with a classifier choosing a different scene. To assess whether weights were significantly different from zero we performed two-sided t-tests across subjects at every pixel location in the visual field, and maps show t-values surpassing a $p < 0.05$ threshold.

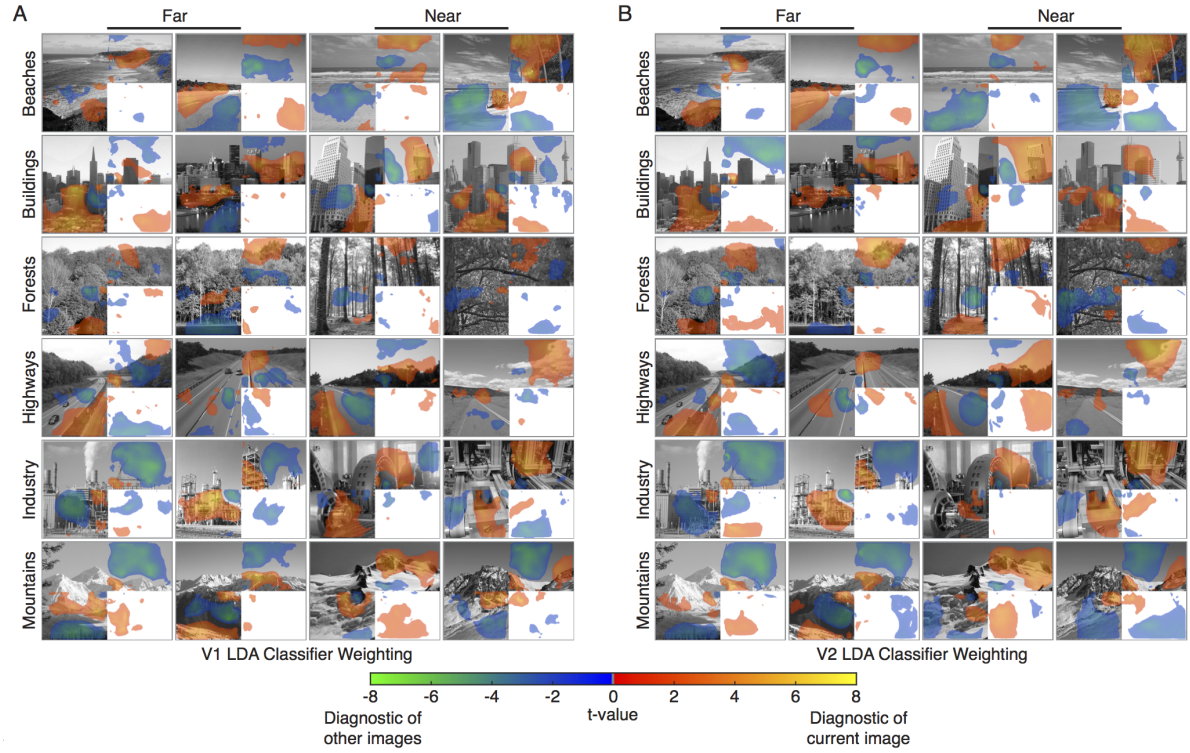

Figure 2.4: Projections of V1 (A) and V2 (B) Linear Discriminant classifier weights into visual space. Voxel classifier weights (the relative contribution of each voxel to an LDA solution) were averaged for significant classifications involving each scene, resulting in one map of discriminatory visual information for each scene in each subject. A two-tailed t-test was conducted across subjects at each pixel location in visual space to obtain t-value maps ($p<0.05$ threshold). Warm colors are visual areas of the scene where voxel activation is indicative of the respective scene. Cool colors are areas where voxel activation is indicative of a different scene. See also Figure A.2.

Classifier information corresponds to visible scene features closely in Non-Occluded areas Figure 2.4. Visual inspection shows that positive weights often match edges or high-contrast scene areas and negative weights match low-contrast areas such as sky or water. In contrast, we did not find that informative areas in the Occluded quadrant are consistently located where objects or contours would continue under the occluder, as might be expected if voxels actively "filled in" objects in occluded portions of scenes. Importantly though, despite relatively small response amplitudes in Occluded areas (Figure A.2), there are retinotopically consistent areas across individuals which are informative to classifier models. These consistent areas do not reside along the Occlusion boundary, indicating that information driving successful decoding is not due to spill-over or edge effects.

Visual inspection of Figure 2.4 indicates that retinotopic patterns are highly consistent between Non-Occluded areas of V1 and V2, consistent with previous reports of responses to natural stimuli (Freeman et al., 2013). Occluded V1 and V2 do not display the same level of consistency. To verify this observation, we tested whether V1 and V2 classifier information was similar enough to enable cross-decoding of scene projection maps in each quadrant by correlating quadrant pixel values from every V1 projection with every V2 projection (Spearman rho) with subjects treated as a random effect. Cross-decoding of V1 and V2 classifier

weights was successful in both Non-Occluded quadrants ($p < 0.001$and $p = 0.02$ for upper-right and lower-left quadrants, respectively), and was unsuccessful in the Occluded quadrant ($p = 0.22$). Therefore, classifier information in Non-Occluded V1 and V2 is similar, but it is not in Occluded V1 and V2. This suggests that retinotopic targets of feedback differ between V1 and V2. This matches anatomical labelling studies which have shown that feedback projections from ventral occipital and parietal areas to V1 and V2 differ retinotopically (Rockland and Hoesen, 1994; Borra and Rockland, 2011).

By projecting RSA voxel weights into visual space, we have provided additional detail to supplement our decoding results and provide an intuitive visualization of features which are used for scene decoding across subjects. In Non-Occluded scene regions, important features include edges and high-contrast objects and are very similar in V1 and V2, matching what is known about response properties in early visual cortex (Freeman et al., 2013; Kay et al., 2008). By design, Occluded regions do not contain useful visual features for decoding, yet projections depict retinotopically consistent locations of information in Occluded responses. These informative areas are not limited to occlusion boundaries and differ between V1 and V2 responses.

2.3 Discussion

We observed activation patterns in Occluded regions of early visual cortex that were informative for determining individual scene and category information about the surrounding images, thus indicating that contextual feedback to early visual cortex is scene-specific, yet exhibits high-level structure. Moreover, we found that responses in Occluded V1 and V2 differed from each other, suggesting that feedback to these two areas have unique retinotopic targets. Our findings are consistent with descriptions of the visual system as a hierarchical inference network, with V1 acting as a high-resolution geometric buffer or blackboard (Lee et al., 1998; Lee and Mumford, 2003). In other words, V1 preserves scene information for reference in calculations where high-resolution image details or spatial precision are required.

The current results and several other studies lend support to the presence of expected information in early visual cortex during perception of absent visual input (Smith and Muckli, 2010; Muckli et al., 2005; Lee and Mumford, 2003; Lee et al., 1998; Muckli et al., 2015; Alink et al., 2008; Sugita, 1999). Still, we are left with the question of how the relatively small high-level effects that we observe in Occluded areas are able to impact cortical processing. We hypothesize that cortical feedback signals can play an important role in determining the output of neurons, and that this mechanism is linked to perception (Takahashi et al., 2016). Evidence from animal models suggests that cortical layer 5 pyramidal cell spiking is virtually unaffected by dendritic stimulation of the apical tuft dendrites, where feedback is largely received. However, cells are tuned to be extremely sensitive to associa-

tive feedback upon receiving feedforward input to their somatic dendrites in a process termed backpropagation-activated Ca2+ spike firing (BAC firing) (Larkum, 2013), where coincident arrival of feedforward and feedback tuft inputs leads to bursts of action potentials. Since the BOLD signal in fMRI is sensitive to energy consumption in both dendritic synaptic processes and spiking activity, as has been shown in primate data (Logothetis, 2008, 2007), we might be detecting dendritic stimulation without feedforward input in Occluded regions of cortex. These synaptically-driven BOLD responses should appear comparatively weak to those caused by the rigorous BAC firing that occurs when feedforward and feedback inputs are integrated. Combined, these points provide one explanation of how relatively small BOLD changes in Occluded regions can be associated with higher-level neuronal processes that significantly affect cortical processing and perception.

These data go beyond the inferential filling-in of expected information in early visual cortex. Our cross-classification results suggest that internal models are based on generalized scene templates that are then constrained by feedforward inputs. The visual system first attends to coarse-scale scene information to quickly estimate input and activate scene templates in memory (Schyns and Oliva, 1994; Oliva and Torralba, 2006). Feedback loops connecting higher areas to early visual cortex may incorporate scene-specific information into the initial template structure. Since fMRI is associated with late components of neural response time courses (Freeman et al., 2013), Occluded responses may display characteristics of initial templates as well as additional scene-specific details. This could explain our ability to classify individual scenes and cross-classify category information; Non-Occluded portions of scenes activate global scene templates that also cover Occluded portions of scenes. Scene-specific additions to these templates would allow for individual scene classification, even between scenes of the same category.

In two separate fMRI experiments, we found that category information is present in feedback to V1 and V2, but did not find evidence to support the presence of depth information in feedback. We also found that the V1 and V2 responses in the lower visual field did not contain depth information in Non-Occluded areas either. However, scene depth was the best tested model for describing scene representations in the Non-Occluded upper visual field. Future studies are needed to fully understand visual field processing biases in V1 and V2 (as well as the rest of the visual hierarchy) and their relationship to scene depth. A recent study showed that several localized early visual cortical patches respond preferentially to objects nearer the viewer (Lescroart et al., 2015). In future, systematically mapping depth properties onto the cortex and manipulating the availability of feedforward image information in such large-scale experimental designs (such as by occlusion) will enable researchers to resolve the contributions of feedforward and feedback signals to these and many other cortical response properties.

Our projection analyses showed that informative areas of Occluded V1 and V2 differ, suggesting that retinotopic targets of feedback to these two areas also differ. These findings are supported by anatomical tracing studies showing that feedback connections to V1 and

V2 terminate in different subsections of these cortical areas (Rockland and Hoesen, 1994; Borra and Rockland, 2011). V1 and V2 receive feedback from many brain areas (Kennedy and Bullier, 1985; Markov et al., 2014), and feedback to each area likely plays a specialized role in predicting aspects of visual scenes. Despite V2's interconnectedness with extrastriate cortex, it is highly dependent on input from V1 due to its role in summarizing natural patterns (Freeman et al., 2013; Serre et al., 2007; Sincich and Horton, 2005). The specific source of contextual information fed to V2 is not known from our results, but our results suggest that V2 might receive more input from Occluded V1 than it is receiving feedback from other areas. In the current study, input to V2 from V1 could be a disparate reconciliation of non-meaningful feedforward input and contextual expectations from feedback (Lee and Mumford, 2003).

Our results are in line with theoretical views explaining the visual system as a hierarchical inference network. Moreover, feedback to Occluded V1 & V2 have unique retinotopic targets. We also found that feedback signals to early visual cortex do not include orientation information. Together these results highlight that a true understanding of the neural computations of early visual cortex will involve understanding more about how and what information is conveyed by feedback.

2.4 Methods

2.4.1 Participants

Twenty-three healthy individuals (N = 18 in main experiment: 12 female, age = 26.45 ± 5.70, mean ± SD; N = 5 in Experiment 2: 2 female, age = 26.50 ± 5.58) with normal or corrected-to-normal vision gave written informed consent to participate in this study, in accordance with the institutional guidelines of the local ethics committee of the College of Science & Engineering at the University of Glasgow (#CSE01127).

2.4.2 Stimuli

Twenty-four real-world scenes from six categories were chosen from a dataset compiled by Walther et al. (2009). Images were displayed in gray-scale (matched for global luminance) on a rear-projection screen using a projector system (1024 x 768 resolution, 60 Hz refresh rate). Stimuli spanned 19.5 x 14.7° of visual angle, and were presented with the lower-right quadrant occluded by a white box (occluded region spanned $\approx$ 9 x 7°). A centralized fixation checkerboard (9 x 9 pixels) marked the center of the scene images. Stimuli in Experiment 2 were identical other than that the occluder was moved to the upper-right quadrant.

2.4.3 Experimental Design

Each of the eight runs consisted of six blocks of eight sequences of stimulation with intervening fixation periods, plus two mapping blocks (total scanning time per run was 804s). Each stimulation sequence lasted 120s, with 12s fixation at the beginning and the end of each series. Stimuli were flashed at a rate of 5Hz in order to maximize the signal-to-noise ratio of the BOLD response (Kay et al., 2008). Each sequence was presented in a pseudo-randomized order where individual images were not shown repeatedly. Over the course of the experiment, each scene was presented 16 times (two times per run). To ensure fixation, we instructed participants to respond via a button press to a temporally random fixation color change. So that participants would attend to the scenes, participants were asked to report the category of the scene being presented during the fixation color change using 6 randomized response buttons.

We used mapping blocks to localize the cortical representation of the occluded region (Muckli et al., 2005). In a block design, subjects viewed contrast-reversing checkerboard stimuli (5Hz) at three visual locations: Target (lower-right quadrant in main experiment, upper-right quadrant in Experiment 2), Surround (of the target), and Control (remaining three quadrants). Each condition was displayed for 12s with a 12s fixation period following, and mapping blocks were randomly inserted between experimental blocks, once per run. We conducted retinotopic mapping (polar-angle and eccentricity) runs separately from the main experiment.

2.4.4 fMRI Acquisition

MRI data were collected at the Centre for Cognitive Neuroimaging, University of Glasgow. T1-weighted anatomical and echo-planar (EPI) images were acquired using a research-dedicated 3T Tim Trio MRI system (Siemens, Erlangen, Germany) with a 32-channel head coil and integrated parallel imaging techniques (IPAT factor: 2). Functional scanning used EPI sequences to acquire partial brain volumes aligned to maximize coverage of early visual areas (18 slices; voxel size: 3mm, isotropic; 0.3mm interslice gap; TR = 1000ms; TE = 30ms; matrix size = 70x64; FOV = 210x192mm). Four runs of the experimental task (804 vol.), one run of retinotopic mapping [session 1: polar angle (808 vol.); session 2: eccentricity (648 vol.)], and a high-resolution anatomical scan (3D MPRAGE, voxel size: 1mm, isotropic) were performed during each of two scanning sessions.

2.4.5 fMRI Data Preprocessing

Functional data for each run were corrected for slice time and 3D motion, temporally filtered (high-pass filter with Fourier basis set [6 cycles], linearly detrended), and spatially normalized to Talairach space using Brain Voyager QX 2.8 (Brain Innovation, Maastricht, Netherlands). No spatial smoothing was performed. These functional data were then overlaid onto

their respective anatomical data in the form of an inflated surface. Retinotopic mapping runs were used to define early visual areas V1 and V2 using linear cross-correlation of eight polar angle conditions.

A general linear model (GLM) with one predictor for each condition (Target > Surround; mapping conditions from experimental runs) was used to define regions of interest (ROI) that responded to the visual target region (lower-right quadrant) and two control regions (upper-right and lower-left quadrants), within V1 and V2. We then performed pRF analyses (Dumoulin and Wandell, 2008) on all ROI voxels, and excluded those voxels whose response profiles were not fully contained (within 2σ of their pRF center) by the respective visual ROI. Lastly, a conjunction of two GLM contrasts (Target > Surround & Target > Control for Occluded ROIs, and Control > Surround & Control > Target for Non-Occluded ROIs) was used to exclude any voxels responding to stimuli presentation outside their respective visual ROI (see Figure 2.1). Time courses from each selected vertex were then extracted independently per run and a GLM was applied to estimate response amplitudes on a single-block basis. The resulting beta weights estimated peak activation for each single block assuming a standard 2γ hemodynamic response function.

2.4.6 Multivariate Pattern Analyses

For SVM classification analyses, a separate regressor modeled each experimental trial. This procedure yielded a pattern of voxel activations for each single trial, and parameter estimates (β values) were obtained for each voxel and then z-scored. A Linear SVM classifier was trained to learn the mapping between a set of all available multivariate observations of brain activity and the particular scene presented, and the classifier was tested on an independent set of test data. Classification analyses were performed using a pairwise multiclass method. Classifier performance was assessed using an n-fold leave-one-run-out cross-validation procedure where models were built on n – 1 runs and were tested on the independent nth run (repeated for the eight runs of the experiment). In analyses of category and depth-based classification, individual scene presentation labels were combined based on these distinctions before training and testing of the SVM classifiers. Significance of individual subject testing was assessed using permutation testing of SVM classifiers. We shuffled data labels in training sets and left testing set labels intact, repeating this procedure 1000 times. This procedure resulted in a null classification model around chance-level and our observed classification value was compared to this distribution to determine the classification significance compared to chance. To determine the group-level distribution of classification performances, we resampled the average observed performances across all subjects 1000 times and report 95% confidence intervals on these bootstrapped values.

Cross-classification analyses were performed similarly to those of our cross-validated classification, but our scene set was split up prior to model training. Training set sizes were 18 and 22 scenes for category and depth analyses, respectively, and testing sets consisted

of the remaining scenes. Due to the large number of possible scene permutations, we conducted 100 iterations of our analyses in each subject. For these analyses, training and testing sets were defined in a pseudo-random manner where each category or depth was evenly represented within both sets. As in our cross-validated classification, training and testing of models occurred on independent data sets using a leave-one-run-out procedure. Since we performed leave-one-run-out cross-validation on each of the 100 training/testing sets, permutation testing for individual-subject classification significance was not feasible, as it would have required 100,000 tests in each ROI, information type and subject. We therefore employed Wilcoxon rank-signed testing to examine individual subject performance. In each of the 100 training/testing sets, we averaged the 8 cross-validated classification performances, resulting in 100 performance values and tested those values against chance-level (See Table S2C). To report group-level performance in these analyses, we averaged over cross-validation folds and then averaged performances over the 100 training/testing sets to get an individual subject performance. These values were bootstrapped to determine 95% confidence intervals.

For RSA, an iterative split-half correlation method (Kravitz et al., 2011; Kriegeskorte et al., 2008) was applied to each subject's data where runs were split into two halves (four runs in each) and concatenated. GLM analyses were then conducted to estimate condition-based responses to each scene in the respective half of the data and repeated for the 35 possible combinations of data splits. Cross-correlation was then used to establish the similarity between the reponse patterns of each pair of scenes. A Fisher transformation was applied to each correlation value before combining correlations into group analyses, and the transformation was reversed for results presentation.

2.4.7 Retinotopic projections

Using each voxel's two-dimensional Gaussian response, as estimated in our pRF analysis, we projected voxel classifier weights and activity patterns into visual space. Voxel pRF functions were multiplied by their respective weights from a Fisher Linear Discriminant Analysis to produce a single visual field map for each scene comparison (each voxel is assigned a weight between -1 and 1). All comparisons involving a scene that had significant classification were averaged to produce a single map for each scene in each subject. We then performed a two-sided t-test across subjects at every pixel location in the visual field in each scene to assess whether weights were significantly different from zero. This procedure was repeated with voxel responses to each scene (mean removed) to produce Figure A.2.

To compare retinotopic patterns in projections, we tested whether V1 and V2 response and classification weights were similar enough to enable cross-decoding of scenes in each quadrant. This procedure involved correlating quadrant pixel values from every V1 scene with every V2 scene (Spearman rho), resulting in a 24x24 scene correlation matrix (similar to an RSA matrix). If the average Fisher-transformed correlation value of the matrix diag-

onal was greater than the average off-diagonal value, then cross-decoding was successful. Subjects were treated as random effects, thus scenes for V1 and V2 were randomly chosen from subjects. This procedure was repeated 1000 times and used to calculate p-values for cross-decoding significance in each quadrant.

Chapter 3

Relating feedback to V1 and V2 to multiple levels of scene information.

Many cortical areas send feedback signals to early visual cortex. Areas with feedback connections to primary visual cortex range in their computational complexity from specializing in low-, mid- and high-level visual features, associative processing, and multisensory information. We utilized data previously described in Chapter 2 to relate responses in non-feedforward stimulated subregions of V1 and V2 to computational models of visual predictions using Representational Similarity Analysis and encoding modeling. We found that multiple levels of predictive information are present in the responses of non-stimulated V1 and V2. Interestingly, foveal portions of V1 and V2 displayed predictions related to low-level visual properties, while more peripheral areas displayed higher-level properties. Our findings suggest that feedback connections from different cortical areas to early visual cortex might be retinotopically organized based on visual properties.

3.1 Introduction

Theories of predictive coding postulate that the brain predicts upcoming perceptual experiences based on internal models. In predictive coding, the brain is essentially an inference engine, aiming to optimize representations of the causes of sensory input (Mumford, 1991; Lee and Mumford, 2003; Clark, 2013). Importantly though, the internal models which are compared to sensory input are not central, but are distributed across the brain. In predictive coding, feedforward and feedback signals are compared at every level of the brain's hierarchical systems (Friston, 2008; Bastos et al., 2012). Each hierarchical level attempts to predict responses at each lower level (Rao and Ballard, 1999). As such, higher areas, such as Inferior Temporal Cortex in the visual system, deal with more abstract information, whereas lower areas, such as Primary Visual Cortex, deal with more concrete data.

However, feedback connections can skip levels in the hierarchy. For instance, V1 receives feedback input from almost every visual processing area (Rockland et al., 1994; Rockland

and Hoesen, 1994; Muckli and Petro, 2013; Petro et al., 2014), and from a number of associative and Parietal areas (Borra and Rockland, 2011). It also receives input from subcortical structures (Markov et al., 2013a, 2014), and from sensory processing areas outside of the visual system, such as auditory cortex (Rockland and Hoesen, 1994; Falchier et al., 2002; Rockland and Ojima, 2003; Wang et al., 2008). The abundance of feedback connections to V1 neurons indicates that in addition to retinal input, V1 neurons are capable of integrating a diverse range of information.

In Chapter 2, we found that feedback to early visual cortex carries high-level scene category information. To do so, we blocked feedforward input to subsections of retinotopic visual cortex using a uniform visual occluder covering one quarter of the visual field (Smith and Muckli, 2010) while participants viewed 24 real-world scenes. Scenes came from six categories (Beaches, Buildings, Forests, Highways, Industry and Mountains), allowing us to decode category information from fMRI responses in Occluded V1 and V2. We also found that individual scenes could be decoded from Occluded responses, which posits an open question: what other types of information are carried by feedback to early visual cortex? In the current study, we aimed to further probe the information characteristics of feedback to early visual cortex by comparing V1 and V2 responses to biologically-inspired computational models ranging in complexity alongside the global scene properties investigated in Chapter 2. We were also interested in whether high-level information in feedback is unique, or whether category information is shared by lower-levels of contextual information.

3.2 Results

We have expanded our analyses of the dataset described in Chapter 2 to compare Occluded V1 and V2 responses to computational models. These biologically-inspired models coarsely spanned the levels of complexity which exist in the visual system along the ventral Occiptial pathway. The Weibull contrast model is inspired by lateral geniculate processing (Scholte et al., 2009; Groen et al., 2013, 2012). The Gist algorithm carries orientation and spatial frequency information, like Gabor filters in V1 (Oliva and Torralba, 2001). The H-MAX model (Layer C2) is tuned to mid-level visual features (combinations of edge segments), and is similar to the tuning of intermediate ventral stream areas such as V4 or posterior IT (Serre et al., 2007). Lastly, global features, Category and Depth, modulate responses in the ventral temporal parahippocampal place area (PPA; Kravitz et al., 2011).

By projecting RSA voxel weights into visual space in Chapter 2, we provided additional detail to supplement our decoding results. We were able to see that in Non-Occluded scene regions, features eliciting higher classifier weighting included edges and high-contrast objects, matching what is known about response properties in early visual cortex (Lee, 1996; Kay et al., 2008; Freeman et al., 2013). By design, Occluded regions did not contain useful visual features for decoding, yet projections depicted retinotopically consistent sources of

information in Occluded responses. However, we did not find evidence that informative areas consistently resemble edge or object continuations into the Occluded regions. The scene features relating to Occluded responses therefore remains an open question. To investigate this question and to expand on results from our decoding and projection analyses, we related scene representations from cortical areas and computational high-level models using RSA and encoding models.

3.2.1 Relating Scene Representations in V1 and V2

Figure 3.1 shows the representational similarities of the six cortical regions with high-level models (Category and Depth) and three popular biologically-inspired computational models: the Weibull contrast model, inspired by lateral geniculate processing (Scholte et al., 2009; Groen et al., 2013, 2012); the Gist algorithm, similar to Gabor filters in V1 (Oliva and Torralba, 2001); and the H-MAX model (Layer C2), matched to tuning properties of intermediate ventral stream areas such as V4 or posterior IT (Serre et al., 2007). We assessed the level of similarity in how each cortical area or model represented the scene set. To evaluate this similarity, we calculated the rank-correlation (Kendall's tau-a) of response differences (Linear Discriminant Contrasts; Walther et al. 2016) in independent data sets. Individual subject correlation values were bootstrapped to produce 95% confidence intervals for each cortical area or model comparison. Grey bands indicate a noise ceiling, which is the comparison of the representational structure in each cortical area with itself in independent data sets (a replicability measure for the representational structure of each area; see Section 3.4.2). Occluded areas had lower replicability levels than Non-Occluded areas, which was anticipated given their much lower response amplitudes and thus smaller first-level distances. All regions had discrimination indices greater than zero ($p \leq 0.005$, one-sample t-test, for each cortical region across subjects), indicating that each cortical region was suitable for RSA (Henriksson et al., 2015).

All areas reliably explained some of the variance of the representational structure of other areas, and retinotopically corresponding V1 and V2 displayed particularly high correlations in Non-Occluded areas (i.e. upper-right and lower-left visual field; see raised blue and red bars in the center and right columns of Figure 3.1). Since correlations were calculated across independent splits of data, similarity cannot be explained by intrinsic fluctuations (Henriksson et al., 2015). High correlation values therefore indicate that Non-Occluded portions of V1 & V2 represent the scenes very similarly. In contrast, Occluded V1 represented scenes more similarly to its Non-Occluded neighboring areas than to its Occluded V2 counterpart ($p < 0.05$, paired-sample t-test, indicated by black lines below bars in Figure 5). Furthermore, we found that Occluded V2 exhibits the opposite pattern; it represents scenes more like Occluded V1 than Non-Occluded areas.

In addition to detailed characterizations of inter-ROI relationships, RSA also allows us to compare cortical ROIs and computational and high-level models. Decoding results were

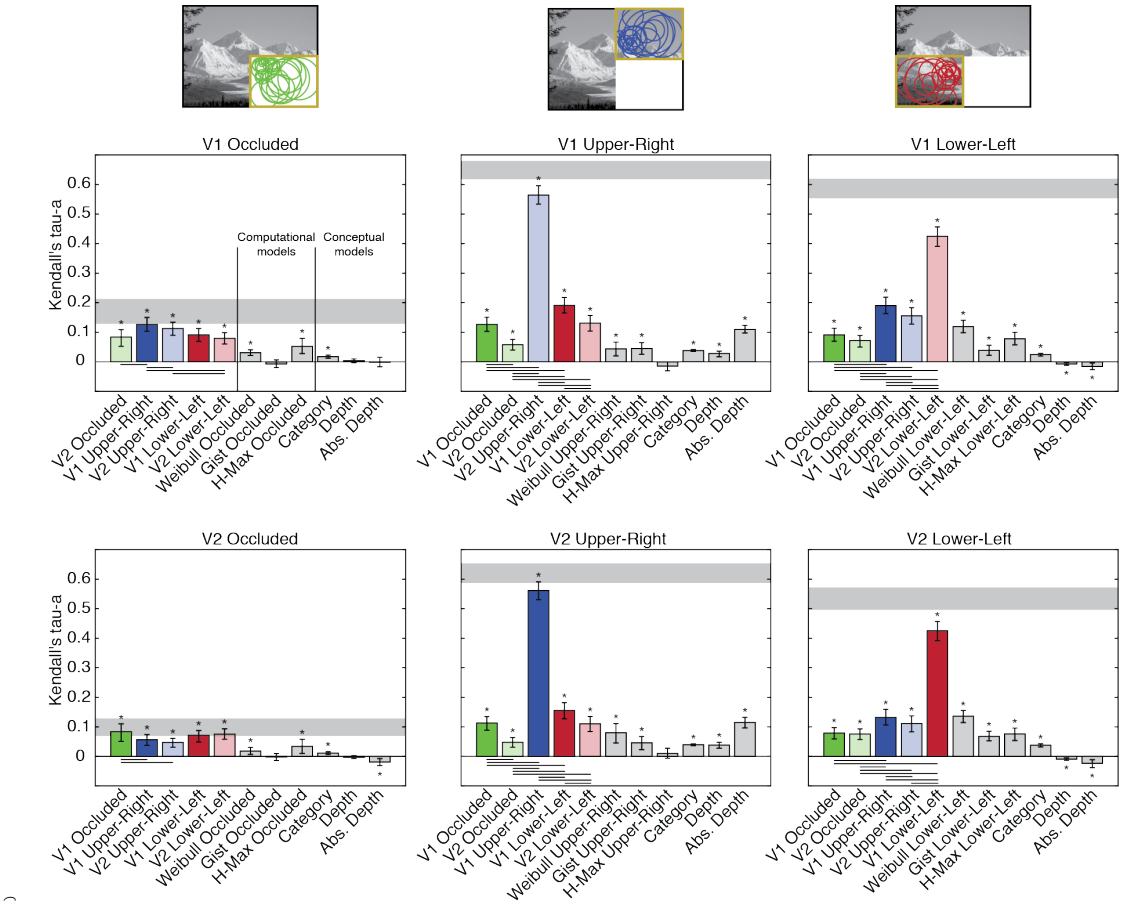

Figure 3.1: Similarity of six cortical regions of interest and computational models measured with Representational Similarity Analysis. Split-quarter (2 runs in each quarter) Linear Discriminant Contrast distances were used to define first-level representational structures. Second-level rank correlations were calculated using split-half Kendall's tau-a between different areas and models (in independent datasets). Individual subject correlation values were bootstrapped to produce 95% confidence intervals. Grey bands indicate a replicability noise ceiling, which is the comparison of the representational structure in each cortical area with itself in independent sets of data. Black lines below cortical regions indicate significant differences in correlations based on a paired-sample t-test ($p<0.05$). See also Figure A.3.

limited to comparisons of discrete scene characteristics (e.g. Beaches vs. Forests, Near vs. Far, etc.), but using RSA, we can compare cortical responses to models with continuous outputs (contrast energy, orientation tuning, etc.). Figure 3.1 makes it easy to see which models explain a portion of each region's representational variance and which are unrelated. Succinctly, the Gist orientation model is unrelated to the representational structure of Occluded areas, and the high-level H-Max model is unrelated to the Upper-Right visual field. All other model comparisons explain a portion of cortical representations in our ROIs. Based on H-Max results, early visual cortical responses contain some high-level information, and this high-level information is invariant to the availability of feedforward information. Further, our results suggest that feedback to V1 and V2 does not contain orientation information, as the Gist model does not explain any variance of Occluded representations.

Importantly, scene depth can be more detailed than just a 'Near' or 'Far' characterization. We therefore conducted a behavioral experiment where 10 subjects rated the scene depths

of each of our 24 scenes in meters (see Appendix A; Figure A.3). With the addition of behavioral depth ratings, we could compare regions to an absolute depth model (Figure 3.1, far right bar in each plot). This model explained more representational variance than any other model for the upper visual field of both V1 and V2 (center column of plots). In the lower visual field, Non-Occluded V1 and V2 were negatively correlated with absolute depth (right column), meaning that these areas had greater response differences to scenes with similar depths than to those with dissimilar depths. In the Occluded regions, V2 also exhibited this negative relationship, while the representational structure of V1 was unrelated to absolute depth (left column).

The negative correlation of Occluded V2 with depth that we observe in RSA is indicative of information about depth, even if it cannot be used for decoding (our discrete Depth model did not exhibit a negative correlation). So, could depth information be fed back to V2, but not fed back to V1? If this is the case, we would expect this effect to be as strong, or perhaps stronger, when the upper visual field is occluded based on the upper-right quadrant's high correlation values with the absolute depth model in our main experiment. To examine this possibility, we analyzed our second fMRI experiment, where the upper-right quadrant was occluded, using RSA. We found that Occluded V1 and V2 were both unrelated to the absolute depth model in these data (not shown). Non-Occluded quadrants (both V1 and V2) had negative correlations with the model, matching results from our main experiment in the lower visual field. Depth information is therefore not consistently fed back to V1 or V2 across the visual field and the lower visual field consistently has a negative relationship with scene depth.

Lastly, high-level scene characteristics are intrinsically linked to low-level visual properties, as evidenced by the impressive capabilities of feedforward deep-learning networks to represent scene categories in their higher layers simply by multiple stages of non-linear statistical summation (LeCun et al., 2015). To investigate whether low-level confounds could explain the high-level effects in our data, we filtered Occluded V1 and V2 RSA distances (as used to create Figure 3.1) by visible low-level properties of scenes (Weibull and Gist model distances from Non-Occluded quadrants) using a GLM. We then ran a separate GLM with Category and both Depth models as predictors on the residuals from this filtering process, which can be thought of as scene decoding that cannot be explained by available low-level visual information. Category was a significant predictor in this analysis ($p < 0.01$ in Occluded V1 and V2), and as expected, Depth was not. Therefore, high-level Category effects in our data cannot be explained by low-level confounds alone.

3.2.2 Voxel-wise Information Encoding

We have investigated and characterized some of the properties of distributed activation patterns in Occluded V1 and V2 using multi-voxel pattern analyses in Chapter 2 and in Section 3.2.1. We were also interested in whether the information portrayed by computational mod-

els and high-level scene characteristics are related to the responses of individual Occluded voxels. We investigated this possibility by mapping Weibull, Gist, H-Max, Category and Depth information onto individual voxel responses using encoding models.

In each voxel that was included in previous analyses, we calculated the responses of the three computational models based on to the portion of each scene spanned by the voxel's pRF. This process resulted in pRF-specific predictors for each voxel as well as two global predictors - Category and the continuous Depth model defined in Section 3.2.1. In the case of Occluded voxels, models were defined using the portion of the scene hidden by the occluder. Models were fitted using partial least squares (PLS) regression to the voxel responses (GLM β weights from previous analyses) to twelve scenes and tested on the responses to the remaining twelve scenes. This process was repeated over 1000 cross-validation folds. We used PLS regression because it is capable of dealing with the number of predictors being much greater than the number of training examples. PLS regression is also capable of separating predictors into covarying components that are related to response variability (Geladi and Kowalski, 1986),

In our analyses, we were particularly interested in the unique contributions of each model. This was because models inevitably share information, and shared information cannot help us determine the importance of each model to overall response predictions. Here, we quantified the unique contributions of each model by calculating their semi-partial correlation statistics, which provide the explanatory power of each model with all other models held constant. Figure 3.2 (left plot) displays the average unique model information for each region of interest. Non-Occluded areas are dominated by low-level information, as the correlations are highest for Weibull and Gist models, and decrease through the mid-level H-Max model and high-level Category and Depth predictors. This finding is very much in line with textbook descriptions of early visual cortex, which describe V1 and V2 responding to low-level features such as oriented edges (Hubel and Wiesel, 1959; Lee, 1996; Olshausen and Field, 2005). In Occluded V1 and V2, mean correlation values for all models were similarly low for all models.

Universally-low information levels in Occluded regions could indicate that very little information is present in the responses of individual Occluded voxels. However, we are able to assess the significance of individual voxels' information levels because we cross-validated our models 1000 times. The middle plot of Figure 3.2 shows the percentage of each region that meets statistical threshold ($p<0.05$) for positive correlation values. Here we see that more Occluded voxels encode Category information than any other type of information. Further, the percentage of Occluded V1 and V2 with significant Category encoding is the same as in Non-Occluded areas.

There is significant encoding of information in individual Occluded voxels, though only in a small proportion of V1 and V2. To understand the information of this population of voxels, we next examined the correlation values of voxels which significantly encode any type of model information (Figure 3.2, right plot). The profiles of Non-Occluded voxels re-

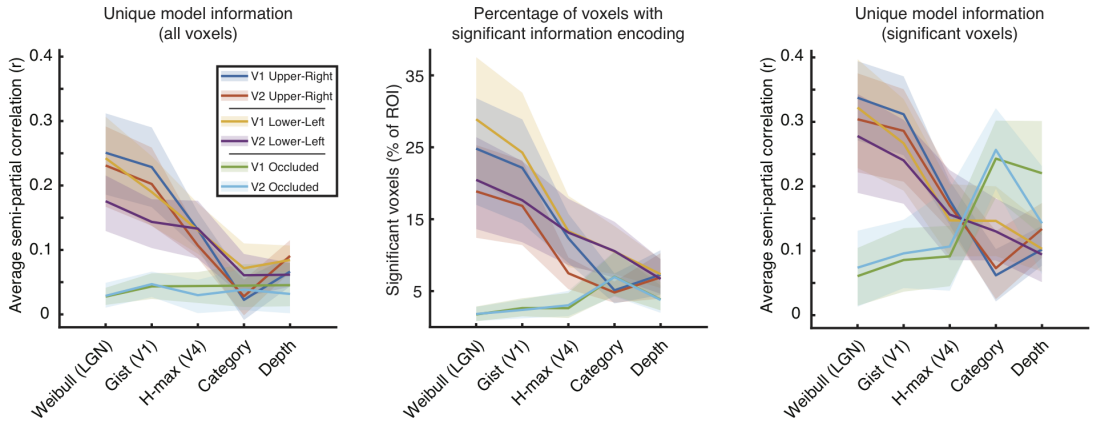

Figure 3.2: Voxelwise unique model information encoding. [Left] Unique model information is displayed as the average of all individual voxel semi-partial correlations of model predictors while all other models are held constant. Solid lines indicate mean correlation and shaded areas indicate 95% confidence intervals (calculated by 1000 bootstrap samples of the mean). [Center] The percentage of voxels in each ROI displaying significant individual voxel encoding for each model ($p < 0.05$; calculated via cross-validated semi-partial correlation 1000 random splits of training and testing scenes). [Right] Unique model information of voxels with significant encoding for any model.

mained similar to those of the full regions, while the information profile of Occluded regions changed dramatically. These Occluded voxels still display less low-level information than Non-Occluded regions, but have more high-level information, even compared to the amount of unique high-level information in Non-Occluded areas.

We have shown that there is increased high-level information in the responses of a small proportion of Occluded voxels, so we next investigated whether informative voxels in different subjects occupy similar retinotopic space. We projected semi-partial correlations from each model into visual space following the same method used in Chapter 2 to project RSA weights. Figure 3.3 displays projections for all voxels and for voxels with statistically significant encoding, matching voxels in the left and right panels of Figure 3.2, respectively. Different areas of Occluded regions encode for model information consistently across subjects ($p<0.01$, FDR corrected). In both V1 and V2, the low-level Gist model (and to a certain extent, the LGN-type Weibull model) occupy more foveal areas of the occluded scenes. In V1, the high-level models (H-Max, Category and Depth) occupy peripheral areas, starting at approximately 5° visual angle from fixation. In V2, high-level model information is not retinotopically localized across subjects, as only a small patch reaches significance for the depth model.

We also asked whether voxels that have significant encoding for model information at an individual voxel level are retinotopically similar across subjects (right plots). In both V1 and V2, this is only true for high-level models (right images below), with V1 patches appearing in a segmented ring containing information about H-Max, Category and Depth at approximately 5° visual angle. V2 patches only appear for Category and Depth and span much of the occluded space. Overall, these results show that high-level information is encoded at a

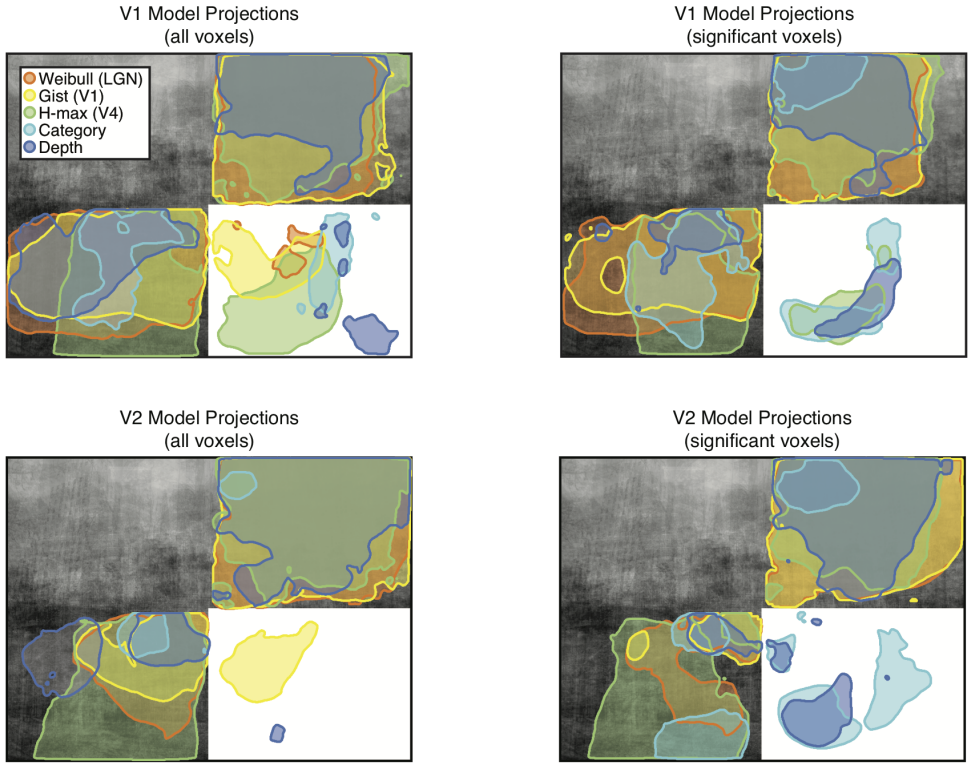

Figure 3.3: Projections of V1 and V2 unique model information encoding into visual space. Fisher-transformed semi-partial correlations for each model were projected into visual space, resulting in one map of unique model information in each subject. A two-tailed t-test was conducted across subjects at each pixel location in visual space to obtain model-specific t-value maps. Maps are overlayed together and colored outlines signify areas reaching a $p<0.05$ threshold (FDR corrected).

statistically significant level by individual voxels not receiving feedforward input. In V1 and V2, these voxels largely reside in peripheral areas, suggesting that the types of information which make up predictions fed back to early visual areas during occluded scene processing are not homogenous across the visual field and that there may be differing roles of feedback between foveal and peripheral areas.

3.3 Discussion

Our data challenge feedforward models of visual processing in early visual cortex; V1 and V2 response patterns to visual stimulation are different than would be predicted by common computational models in Occluded regions. Models would predict identical responses to the white occluder because they do not account for contributions from contextual feedback. However, we observed activation patterns in Occluded regions that were informative for determining individual scene and category information about the surrounding images (Chapter 2), thus indicating that contextual feedback to early visual cortex is scene-specific, yet exhibits high-level structure. In the current chapter, we found evidence that multiple levels of predictive information about occluded portions of scenes are present in the responses of

Occluded V1 and V2. Such findings are consistent with descriptions of the visual system as a hierarchical inference network, with V1 acting as a high-resolution geometric buffer or blackboard (Lee et al., 1998; Lee and Mumford, 2003). In other words, V1 preserves scene information for reference in calculations where high-resolution image details or spatial precision are required. At first glance, this description seems incongruent with the idea that V1 can depict visual information that is missing or hidden from view, such as that masked by the occluder in our study. However, feedback to V1 depicts detailed internal models that share information characteristics with expected or predictable feedforward input (Mumford, 1992). Due to the large number of areas that send feedback to V1, it is logical that these internal models span multiple levels of visual complexity.

We projected classifier weights into visual space in Chapter 2, showing that informative Occluded voxels occupy retinotopically similar areas of the visual field across subjects. These areas did not correlate with edges which extend into the occluded area. Anatomical tracing studies have shown that feedback terminations do not correlate with the compartmental organization in V1 and V2 and suggest that feedback might play a distinct functional role from feedforward processing in early visual cortex (Rockland and Hoesen, 1994; Borra and Rockland, 2011). Our RSA results support the possibility that in the absence of meaningful feedforward information, V1 and V2 responses no longer correlate with orientation properties of the occluded portions of scenes. This finding might suggest that edge detection is only a feedforward property of early visual areas, and not of information carried by feedback. However, results from Occluded V1 and V2 encoding models showed that orientation information was an important aspect of Occluded predictions in the fovea.

What could the source of these discrepant results be? One possibility lies in the retinotopic organization of the Gist model compared to early visual cortex. Due to cortical magnification, the density and resolution of cortical scene sampling is much higher in the fovea than in the periphery (Duncan and Boynton, 2003). In the Gist model, this is not the case - scenes are split into a 4×4 grid, where orientation and spatial frequency are summarized within each of the 16 evenly-spaced receptive fields. Encoding models in our study calculated orientation and spatial frequency information using the Gist model within each voxel's pRF, thus more closely matching the sampling of V1 and V2. These results highlight the importance of retinotopic structure when assessing model similarity to early visual cortical responses.

By measuring the representational similarity between different cortical regions and computational models, our analyses aimed to characterize properties of distributed response patterns associated with feedback to V1 and V2. We found that the relationships between Occluded V1, Occluded V2, and their Non-Occluded counterparts do not match. This result refined our conclusions about differing projections of classification weights into visual space in Occluded V1 and V2. Rather than simply not being similar, which is the only conclusion we could make from the negative result in Chapter 2, we can now see that Occluded V1 is more related to its Non-Occluded counterpart than Occluded V2. V1 therefore might receive

more contextual input from the rest of the visual field than V2 does, thus adding an explanation for the higher SVM decoding results in V1 compared to V2 in Chapter 2. Occluded V2 was most similar in its scene representations to Occluded V1, suggesting that V2 might receive more input from a subsection of V1 that is only receiving feedback signals than it is receiving feedback from other areas.

The computational models that we compared to cortical representations are inspired by biological visual systems, but they process scene statistics in a strictly feedforward manner. High-level feature detection has enabled deep learning networks to achieve incredibly high performance on a number of natural signal processing tasks including visual object and speech recognition (LeCun et al., 2015). Still, such networks predominately use feedforward architectures, and would therefore perform sub-optimally when presented with partially occluded visual scenes because they are unable to recognize occluded portions of the visual field as missing information. Occluded areas would be integrated identically to the rest of the scene, but without contributing any useful information. Recent work has shown that by including feedback and lateral connections, convolutional networks are able to outperform those with only feedforward connections on tasks involving occlusion (Spoerer et al., 2017). Additionally, networks utilizing synchronization (a form of feedback) by allowing complex weights between network nodes can disentangle occluded shapes (Reichert and Serre, 2013). We have shown that Occluded V1 and V2 responses contain predictive information about missing feedforward input. Combined with these recent neural network results, we suggest that models of early visual cortex would benefit from the addition of feedback components. Further development of such networks will explain aspects of cortical responses not yet captured by current computational models, and may provide insight to mechanisms of feedback processing in early visual cortex.

3.4 Methods

3.4.1 fMRI Data

Analyses were performed on the 18-subject fMRI dataset described in full in Section 2.4. Functional scanning was aligned to maximize coverage of early visual areas (18 slices; voxel size: 3mm, isotropic; 0.3mm interslice gap; TR = 1000ms; TE = 30ms; matrix size = 70x64; FOV = 210x192mm). Eight runs consisted of six blocks of eight sequences of stimulation with intervening fixation periods, plus two mapping blocks (total scanning time per run was 804s). Each stimulation sequence lasted 120s, with 12s fixation at the beginning and the end of each series. Stimuli included twenty-four real-world scenes from six categories were chosen from a dataset compiled by Walther et al. (2009) that were presented with the lower-right quadrant occluded by a white box.

3.4.2 Representational Similarity Analysis

An iterative split-half correlation method (Kravitz et al., 2011; Kriegeskorte et al., 2008) was applied to each subject's data where runs were split into two halves (four runs in each) and concatenated. GLM analyses were then conducted to estimate condition-based responses to each scene in the respective half of the data and repeated for the 35 possible combinations of data splits. Cross-correlation was then used to establish the similarity between the reponse patterns of each pair of scenes. A Fisher transformation was applied to each correlation value before combining correlations into group analyses, and the transformation was reversed for results presentation.

We assessed second-order correlations of dissimilarity matrices for Occluded and Non-Occluded V1 and V2 ROIs, idealized Category and Depth models, an Absolute Depth model, and three computational models: Weibull (Groen et al., 2013, 2012), Gist (Oliva and Torralba, 2006), and H-MAX C2 (Serre et al., 2007). Kendall's tau-a was calculated in a cross-validated fashion where all comparisons were made between independent sets of data (split-quarter Linear Discriminant Contrast (Walther et al., 2016) distances in first-level and split-half Kendall's tau-a in second-level). This process avoids the increased correlations associated with intrinsic fluctuations reported in Henriksson et al. (2015). Kendall's Tau-a was used because we were comparing response patterns to models that predict rank-ties between scenes (Nili et al., 2014). Values from each possible data fold were averaged within subjects, as folds are not independent from each other (Nili et al., 2014). Individual subject correlation values were then bootstrapped to produce 95% confidence intervals. Grey bands indicate a noise ceiling, which is the comparison of the representational structure in each cortical area with itself in independent sets of data. This is computed in the same way as other comparisons and can be thought of a reliability measure for the representational structure of each area. To confirm the suitability of these data for second-level RSA, a cortical region discrimination index was calculated for each area (Henriksson et al., 2015). This index was calculated by subtracting the off-diagonal values involving the cortical area from the replicability value for each area, calculated for each subject individually.

3.4.3 Encoding Models

In each voxel, we calculated the responses of three computational models (Weibull, Gist, and H-MAX; see individual sections for detail) based on to the portion of each scene spanned by the voxel's pRF. This process resulted in pRF-specific predictors for each voxel as well as two global predictors - Category and the continuous Depth model defined in Section 3.2.1. In the case of Occluded voxels, models were defined using the portion of the scene hidden by the occluder. In total, each model consisted of 2041 voxel-specific predictors (2 Weibull, 32 Gist, 2000 H-MAX, 6 Category, and 1 Depth predictor).

Models were fitted using partial least squares (PLS) regression (via the SIM PLS algo-

rithm; de Jong, 1993) to voxel responses (GLM β weights from previous analyses) to twelve scenes and tested on the responses to the remaining twelve scenes. PLS regression was used because it is capable of dealing with the number of predictors being much greater than the number of training examples by separating predictors into covarying components that are related to response variability (Geladi and Kowalski, 1986). This process was repeated over 1000 pseudo-randomly chosen cross-validation folds where training and testing sets contained even numbers of category exemplars (the only discrete variable in the model).

In each fold of the data, we calculated semi-partial correlation statistics for each individual model:

$$r_{1(2.3)} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{1 - r_{23}^2}} \tag{3.1}$$

Where r_{12} is the Pearson correlation of testing data and predicted responses based on the model of interest, r_{13} is the correlation of testing data and predicted responses based on the remaining four models, and r_{23} is the correlation between predicted responses from the model of interest and the predicted responses from the remaining four models. The use of semi-partial correlations, as opposed to squared semi-partial correlations which can be used for variance decomposition, was that $r_{1(2.3)}$ values are zero-centered, meaning that we could perform significance testing on the 1000 cross-validation folds to see if correlations were greater than zero. Further, correlation values are t-distributed after a Fisher Z transformation. This allowed us to validly calculate mean values within regions of interest and to perform t-testing of these values in projection analyses.

3.4.4 Weibull Model

The Weibull image contrast model measures the distribution of contrast values for an image and seeks to emulate the X and Y cells in the Lateral Geniculate. It has a two-dimensional output. We used the Weibull image contrast model outlined in Groen et al. (2012, 2013) with the field of view for estimation of the beta and gamma parameters set to 1.5 and 5 degrees, respectively.

For encoding models, Par and Mag components of the model (X and Y cell models, respectively) were calculated as maps. Voxelwise contrast energy (the first component of the model) was estimated as the weighted mean of Par maps based on each voxel's pRF:

$$\bar{x} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} \tag{3.2}$$

Where x represents the Par map values and w represents the values of each voxel's two-dimensional pRF function. Spatial coherence (the second component of the model) was

estimated as the weighted variance of the Mag maps based on each voxel's pRF:

$$\hat{\sigma}_w^2 = \frac{\sum_{i=1}^{n} w_i (x_i - \bar{x})^2}{\sum_{i=1}^{n} w_i} \tag{3.3}$$

Where x represents the Mag map values, $\bar{x}$ is the weighted mean of the Mag values based on the pRF, and w represents pRF values.

3.4.5 Gist Model

The Gist algorithm (Oliva and Torralba, 2001) measures the distribution of oriented bandpass Gabor filter responses in localized portions of images. Our model consisted of 16 locations (4 x 4 grid), 8 orientations, and 4 spatial frequencies. This model had a 512-dimensional output.

For encoding models, we calculated maps for each feature (8 orientations and 4 spatial frequencies), resulting in 32 maps. Voxelwise Gist features were then calculated as the weighted mean of the maps based on each voxel's pRF (Equation 3.2).

3.4.6 H-MAX Model

The H-MAX model is a hierarchical model which gradually combines visual features of higher complexity. Here, we used the output of its fourth sequential stage, C2. The first two stages (S1 and C1) correspond to the simple and complex cells or early visual cortex. Stages S2 and C2 use the same pooling mechanisms as S1 and C1, but pool from the C1 stage and respond most strongly to a particular prototype input pattern. Prototypes were learned from a database of natural images outside of this study (Serre et al., 2007). The output of this model had 2000 dimensions.

In encoding models, S2 layer maps were defined and the maximization operation used to calculate C2 layer features was limited to the region of the image within each voxel pRF (defined as 2σ from pRF center). This resulted in 2000 predictor channels per voxel.

3.4.7 Model redundancy measurements

Due to the large number of model predictors, before fitting model data to voxel responses, we tested the level of redundancy within Gist and H-MAX models. This analysis was performed by calculating Gist and H-MAX features in randomly chosen voxel pRFs and images, 100,000 times. Voxels were chosen from any subject and ROI in our analyses, and images were chosen from the 131,000 image SUN database (Xiao et al., 2010). We then performed principal component analysis to determine what dimensionality was required to describe these models.

We found that the Gist model was moderately redundant, as approximately 96% of the model's variance could be described by 15 components (32 original predictors). The H-MAX model was highly redundant, with 96% of its variance being explained by 14 components (2000 original predictors). Based on these findings, we included both models in voxel response models since PLS regression works to combine these highly covarying separate predictors.

3.4.8 Absolute Depth Model

To quantify the depths of our scenes, 10 subjects were asked to rate in meters the absolute depth for each of our 24 scenes. To ensure we were probing the perceptual depth of each scene, participants were given minimal guidance on the definition of the term "scene depth." Depth ratings were converted to a log10 scale, bootstrapped via 1000 samples of the mean, and a normal distribution was fit to the resampling histogram for each scene. This produced probability distributions for each scene's depth, similar to depth ratings in Torralba and Oliva (2002).

3.4.9 Retinotopic projections

Using each voxel's two-dimensional Gaussian response, as estimated in our pRF analysis, we projected semi-partial correlation statistics into visual space. Voxel pRF functions were multiplied by their respective correlation values for each model and summed across voxel pRFs within each subject's data. We then performed a two-sided t-test across subjects at every pixel location in the visual field to assess whether correlations were significantly different from zero. Contours define areas meeting statistical threshold (FDR corrected to $q<0.05$). This procedure was repeated for each model.

Chapter 4

Cortical feedback to superficial layers of V1 contains predictive scene information.

Cortical neurons receive a combination of feedforward and feedback input, and these input streams arrive in different cortical layers. Many experiments have progressed our understanding of the feedforward features that modulate early sensory areas, but relatively little is known about the feature space that drives cortical feedback channels. We blocked feedforward input to subsections of retinotopic visual cortex by occluding one quarter of the visual field while participants viewed 384 real-world scenes and recorded V1 responses using high-resolution 7T fMRI ($0.8mm^3$). We provide evidence that V1 responses exhibit predictive and contextual response properties in addition to feedforward orientation and spatial frequency properties typically associated with V1 responses and that these predictive and contextual responses are primarily associated with superficial layers of cortex. Our findings suggest that feedback connections terminating in superficial layers provide V1 neurons with contextual information not available via localized feedforward input.

4.1 Introduction

Neuronal activity in nearly all cortical regions depends on a combination of feedforward and feedback signals (Bastos et al., 2012; Gilbert and Li, 2013; Markov et al., 2013b). In early visual cortex, neurons receive feedforward sensory signals from the retina and receive modulatory predictive signals through lateral connections and cortical feedback (Smith and Muckli, 2010; Phillips et al., 2015). Only 5% of input to V1 neurons corresponds to feedforward retinal input, and only 20% of V1 neuronal activity can be explained by retinal input (Carandini, 2005; Muckli and Petro, 2013). V1 responses must therefore be highly influenced by feedback and lateral connections. Substantial progress has been made mapping the feedforward features that modulate early visual cortex, but measuring the feature space that drives cortical feedback channels presents a significant challenge for systems neuroscience.

By integrating feedforward and feedback signals, the brain can combine sensory retinal

input with context and knowledge from internal models of the world (Friston, 2008; Clark, 2013). This integration is remarkably adaptive and flexible because it occurs at the level of local microcircuits throughout sensory and associative cortex rather than at a more centralized brain location (Mumford, 1992; Spratling, 2008; Bastos et al., 2012). Feedforward and feedback signals originate and terminate in different cortical layers (Markov and Kennedy, 2013), with feedforward connections coming to middle layers of cortex and feedback coming to superficial and deep layers of cortex. Human neuroimaging experiments have assessed feedback responses in different cortical depths using high-resolution ultra high-field MRI (Muckli et al., 2015; Kok et al., 2016). However, the information content of these feedback responses remains an open question.

In Chapter 3, we found that multiple levels of predictive information are present in the responses of non-stimulated V1 and V2. To do so, we blocked feedforward input to subsections of retinotopic visual cortex using a uniform visual occluder covering one quarter of the visual field (Smith and Muckli, 2010) while participants viewed real-world scenes. We related responses in Occluded V1 and V2 to models of visual predictions to uncover what types of information were present in Occluded responses. In the current study, we extended this question and asked whether feedforward sensory input and predictive feedback signals could be read out from different layers of cortex, and what types of information these two inputs to early visual cortical neurons contain. By utilizing an occlusion paradigm in combination with recordings from high-resolution fMRI, we compared depth-specific V1 responses to sensory- and prediction-based computational models and to high-level scene category information.

4.2 Results

We exploited the retinotopic organization of visual cortex and blocked meaningful sensory input to a portion of primary visual cortex by occluding one quarter of visual real-world scenes (Smith and Muckli, 2010). We recorded brain activity using high-resolution 7T fMRI, allowing us to assign functional voxels to six different cortical depth layers (Muckli et al., 2015). To uncover the information properties of different cortical depths, we mapped sensory and predictive information processing onto voxel responses to 192 Occluded and 192 Non-Occluded scenes using encoding models. In each voxel, we defined voxel-specific predictors relating to feedforward and feedback information streams. Feedforward models were defined as model responses to scenes as they were *presented*. In other words, if a V1 neuron's receptive field was located in the lower-right quadrant of the scenes, it would receive meaningful sensory input during Non-Occluded scene presentations, but would not receive meaningful input during Occluded scenes. For feedforward models, we calculated the responses of two low-level computational models (Weibull and Gist) based on the portion of each scene spanned by the voxel's pRF exactly as scenes were presented (Figure 4.1). The

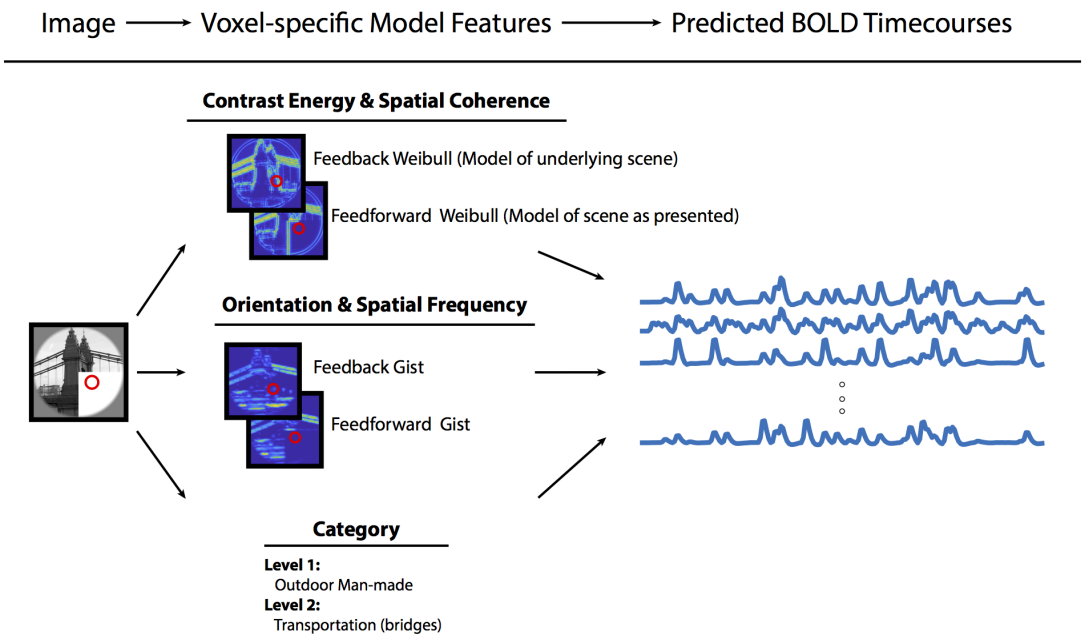

Figure 4.1: Creation of voxel-specific feature timecourses. The process of creating feature-based predicted timecourses is shown for one voxel (the pRF of this example voxel is shown in red). Each stimulus image was decomposed into feedforward and feedback Weibull and Gist feature maps. Feedforward maps were based on the image as it was presented, with the occlusion in place. Feedback maps treated the image as if it had not been occluded, and therefore feedback features were based on possible predictive voxel responses. Additionally, a high-level category model based on the SUN database hierarchy was included (Xiao et al., 2010). Feature responses were used to create predicted timecourses, which were convolved with a hemodynamic response function. These feature timecourses were used for encoding models.

Weibull model is based on contrast energy and spatial coherence information (Groen et al., 2012, 2013), and the Gist model is orientation and spatial frequency-based (Oliva and Torralba, 2001, 2006).

Feedback models aimed to uncover *predictive* voxel responses. These models treated each scene as if it was Non-Occluded even if the scene was presented with occlusion. We calculated pRF-specific responses of the Weibull and Gist models based on Non-Occluded versions of each scene. Additionally, a high-level Category model was also included as a contextual feedback model. Since the occluder only ever appeared in lower-right quadrant of scenes, only voxels with pRFs in that area included predictive feedback (Weibull and Gist) models. In Non-Occluded areas, predictors for feedforward and feedback models were identical. Models were fitted to voxel time courses and tested on left out data. To determine the amount of unique information each model contributed to voxel predictions, we calculated semi-partial correlation statistics for each individual model.

4.2.1 Depth-dependent information encoding

Figure 4.2 displays average voxelwise unique model information encoding in six cortical depth ROIs of three subjects. These values have been corrected for fMRI amplitude increases toward the cortical surface in gradient echo images by normalizing by the mean image re-

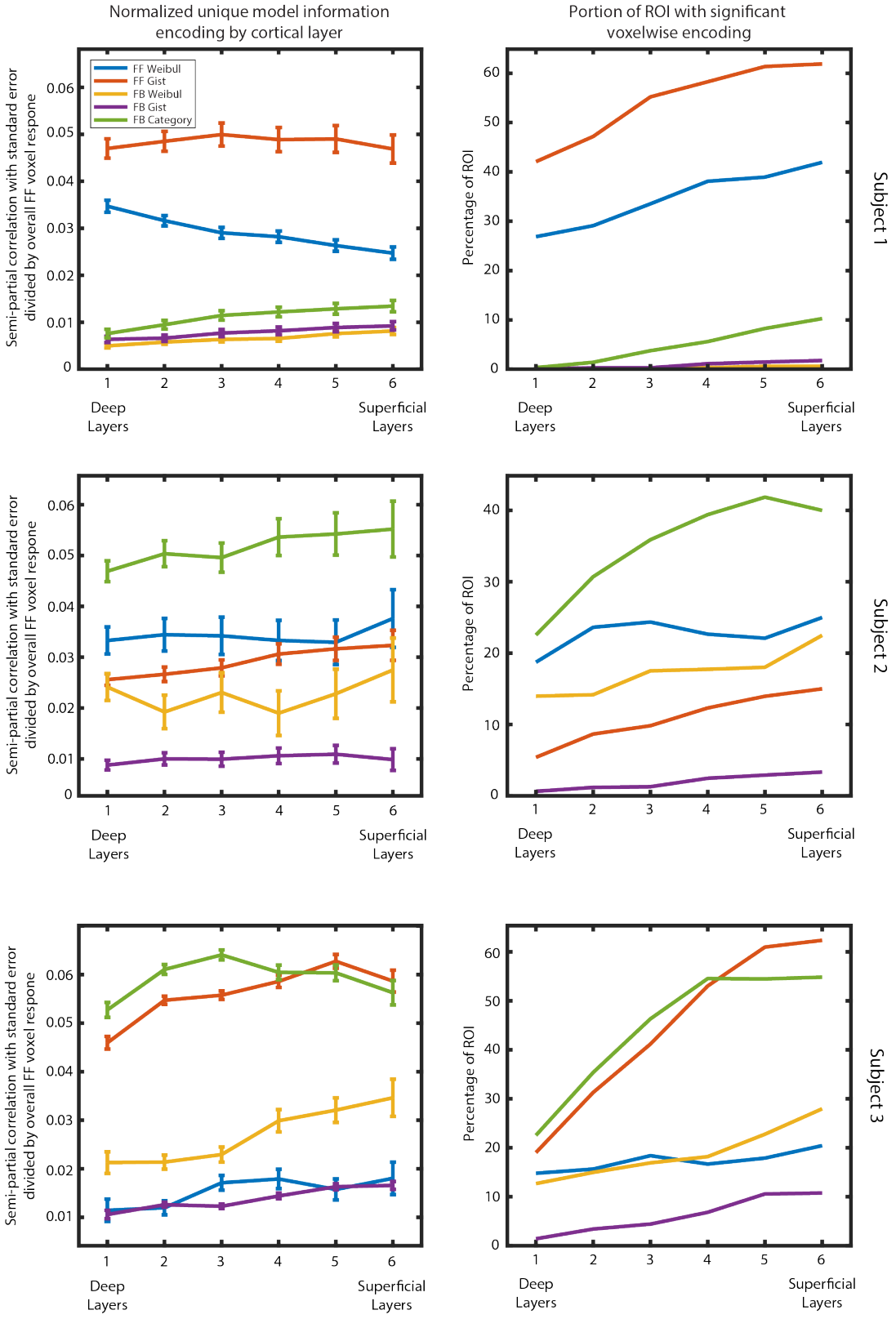

Figure 4.2: Voxelwise unique model information encoding by cortical depth. [Left Column] Unique model information is displayed as the average of all voxel-model semi-partial correlations with standard error at 6 cortical depths [10%, 26%, 42%, 58%, 74%, 90%], with Layer 1 being labeled as the deepest layer and Layer 6 as the most superficial. Layers are normalized by their mean image response to all Non-Occluded images. [Right Column] The percentage of each ROI displaying significant voxelwise encoding for each model ($p < 0.05$; calculated via 1000 cross-validation splits of training and testing data).

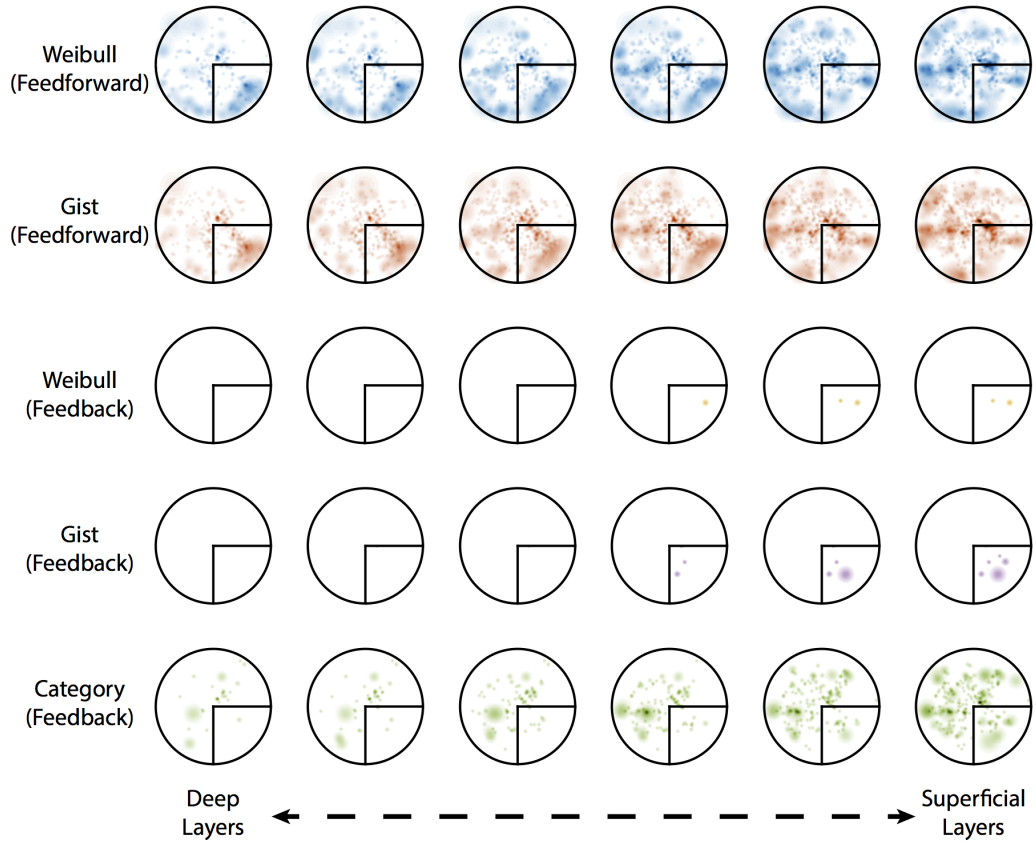

Figure 4.3: Depth-specific projections of V1 voxels with significant model information encoding into visual space. Voxels from Figure 4.2, bottom row, are projected into visual space by summing their 2D Gaussian pRF functions. Feedback Weibull and Gist models (rows 3 & 4) only have Occluded components because these models do not exist for Non-Occluded scene presentations.

sponse to all Non-Occluded images within that layer (uncorrected values are displayed in Figure B.1). In Subject 1, all layers show more information for feedforward models than for feedback models, and Gist is the more informative of the two feedforward models. This makes sense given V1's prominent orientation and spatial frequency sensitivity (Hubel and Wiesel, 1959; Kay et al., 2008). In Subjects 2 and 3, Category contributes the most unique information of any model. The feedforward Gist model contributes equal unique information in Subject 3 and much less information in Subject 2. Importantly though, all feedback models display unique information, and the amount of feedback information largely increases from deep to superficial layers, where many feedback connections to V1 terminate (Douglas and Martin, 1991; Felleman and Van Essen, 1991; Petro and Muckli, 2017).

The right column of Figure 4.2 shows the percentage of voxels in each layer with significant model encoding. All models have increasing percentages of significant voxel encoding from deep to superficial layers (these plots were not normalized by response amplitude), and notably, there is significant encoding for all feedback models in the most superficial layer of the Occluded area. We projected voxels with significant encoding from Subject 1 (the subject with most bias toward feedback information encoding) into visual space by summing their pRFs (Figure 4.3). Voxels which code feedforward information are largely

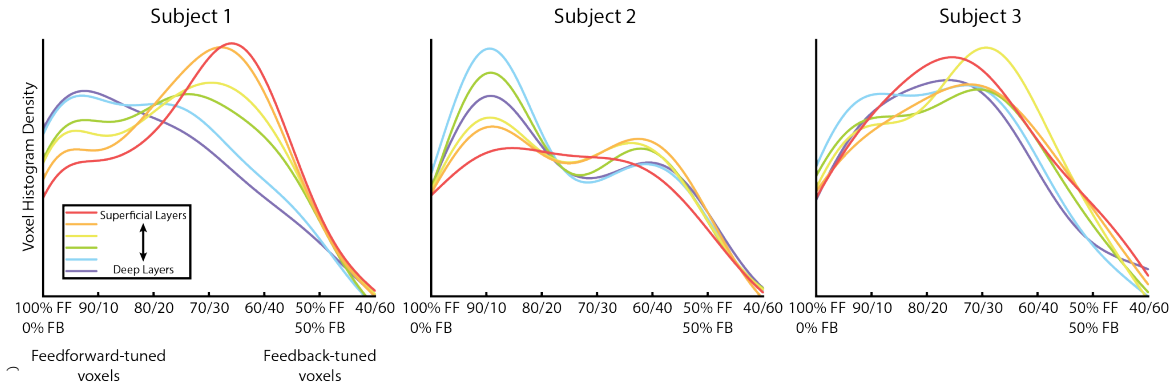

Figure 4.4: Depth-specific feedforward and feedback voxel tuning. Voxel tuning density plots are shown for each cortical depth separately. Tuning was calculated by contrasting the unique information encoding of feedforward models to that of feedback models for each voxel. Probability density functions were defined for each cortical depth ROI using kernel density estimation.

homogeneous within the visual field and increase through layers, as observed in Figure 4.2. Feedback Weibull and Gist models only have Occluded components, but importantly, voxels with significant encoding for these models do not have pRFs at the edge of the occluder, thus increasing confidence that voxels receive predictive signals from feedback. Category information encoding is also generally homogeneous across the visual field, and these voxels cover much of the Occluded area.

4.2.2 Individual voxel tuning to feedforward and feedback signals

We have shown that the amount of unique feedforward and feedback information differs across layers, but it is unclear how feedforward and feedback information is balanced in individual voxels. To investigate individual voxel tuning to sensory and predictive signals, we converted the previously calculated semi-partial correlations to compositional values by dividing each model's value by the sum of all correlation values within each voxel. We were then able to sum the two feedforward models to get a measure of the total tuning toward sensory input. The same was done with feedback models to get tuning toward contextual and predictive signals. Figure 4.4 shows the voxel tuning histograms for each cortical depth. In Subject 1's Occluded area, voxels in deep layers are more likely to be tuned toward feedforward models, with a peak probability of more than 90% feedforward tuning. Voxels in superficial layers are tuned more toward feedback models, with approximately 33% feedback tuning. Subject 2 also shows a trend of feedforward tuning in deep layers which gradually become more tuned to feedback signals in superficial layers. However, this effect is not as strong in Subject 2 as it is in Subject 1. Subject 3 does not show any clear differences in layer tuning.

One potential source of these differences is the size of each subject's Occluded subsection of V1. Subjects had 1502, 791, and 335 voxels in their Occluded cortical patches, respectively. Additionally, Subject 3 had substantially more vasculature in the occluded patch than

the other two subjects, resulting in substantially smaller superficial layer ROIs than other layers (73 voxels in the most superficial layer). These differences suggest that a larger sample size is required to fully understand differences in feedforward and feedback tuning between cortical layers.

4.3 Discussion

Our current understanding of cortical feedforward and feedback connectivity and function is still in its infancy (Markov et al., 2014). Anatomical studies have made great strides in mapping feedback connectivity patterns to different layers of cortex (Felleman and Van Essen, 1991; Rockland et al., 1994; Rockland and Ojima, 2003; Markov and Kennedy, 2013; Markov et al., 2013b), but how these patterns relate to brain function remains a largely open question. A promising theory points to the segregated arrival of feedback and feedforward inputs to the apical and basal dendrites of cortical layer 5 pyramidal neurons (Larkum, 2013). This theory posits that in addition to feedforward action on basal dendrites in middle to deep layers of cortex, pyramidal neuron activity is affected by feedback inputs arriving to apical tuft dendrites in superficial layer 1, which trigger Ca^{2+} spikes. Through a mechanism known as backpropagation-activated Ca^{2+} spike firing (BAC firing), these Ca^{2+} spikes can convert a single somatic output spike into a 10 ms burst containing 2–4 spikes (Larkum et al., 1999), meaning that feedback inputs might have a substantially greater role in determining the firing of pyramidal neurons than previously understood (Larkum et al., 2009).

High-resolution fMRI provides a valuable tool for the examination of feedback in human cortex. fMRI is sensitive to dendritic energy consumption that might not translate to spiking activity and thus not appear in cellular recordings (Petro and Muckli, 2017). Further, it provides sub-millimeter resolution of human cortex at the level of layers (Olman et al., 2012) and columns (Yacoub et al., 2008) - important features to separate feedforward and feedback processes occurring in different layers. In the current study, we utilized the spatial resolution of high-resolution 7T fMRI to record cortical depth-specific V1 responses while we blocked feedforward input to subsections of retinotopic visual cortex by occluding one quarter of the visual field (Smith and Muckli, 2010; Muckli et al., 2015). Participants viewed 384 real-world scenes, allowing us to build models of individual voxels based on feedforward signals of local sensory input as well as predictive and high-level feedback signals. We have shown in two of three subjects that V1 responses exhibit predictive and high-level response properties in addition to feedforward orientation and spatial frequency properties typically associated with V1 responses. Furthermore, predictive and high-level responses were associated primarily with superficial layers of cortex, supporting the known anatomy of connections believed to deliver predictive signals (Felleman and Van Essen, 1991; Friston, 2008; Bastos et al., 2012).

An important factor in analyzing high-resolution fMRI data is the contribution of vascu-

lature, particularly when examining differences between cortical layers. Data from our third subject had substantial vasculature in their Occluded patch of cortex, and this adds a potential source of noise to analyses. Interestingly, this topic has been very recently addressed in a critical assessment of 7T methods (Kay et al., 2018), indicating that dealing with vasculature poses a significant challenge in 7T fMRI studies. Careful examination of vascular patterns is clearly important for intensive 7T fMRI studies, and in studies of occluded cortical areas, it will be useful to map out these patches of cortex prior to collection of large-scale data.

While it is appealing to imagine that the mere use of layer-specific measures of human brain activity allows access to feedback signals, we have shown that studying feedback requires experimental manipulations aimed at separating feedforward and feedback signals. Figure 4.4 shows that the tuning of voxels toward feedforward or feedback signals is different between cortical layers of V1, but importantly it also shows that this is only the case for Occluded voxels. In Non-Occluded voxels, all cortical layers have similar tuning. The mechanism underlying this observation might be related to BAC firing. Larkum (2013) explains that the conceptual importance of this mechanism is that pyramidal neurons are able to detect coincident input to basal and apical dendritic regions. This detective capability provides the cortex with an associative mechanism to combine feedforward and feedback information. In our data, voxels in Non-Occluded V1 have matching sensory and predictive signals, and could therefore have substantial energy consumption in middle and deep layers to sustain bursting activity. These signals are inseparable from feedback responses due to matching dendritic energy consumption in superficial layers. However, in Occluded voxels, the same apical dendritic energy consumption is not accompanied by BAC firing, thus allowing us to isolate information associated with feedback.

We have shown that occlusion of real-world scenes triggers feedback signals to early visual cortex about scene information not available via localized feedforward input, and that such signals can be detected and described using cortical depth-specific fMRI. These findings progress our understanding of intracortical interactions, but do not fully describe the modulatory actions of feedback on V1. A brief list of cognitive functions that feedback to V1 has also been associated with includes motion perception and action (Alink et al., 2010; Vetter et al., 2013; Ban et al., 2013; Edwards et al., 2017), visual imagery and memory (Vetter et al., 2014; Naselaris et al., 2015; Hindy et al., 2016), and reward (Tan, 2009). Understanding more about what cortical layers are associated with these functions of feedback will help uncover new information about the underlying circuits and the role V1 plays in higher-level brain functioning.

Our analysis of voxel tuning toward feedforward or feedback signals provides important information on the involvement of different cortical layers in processing sensory and predictive signals. However, it also begs the question of how specific information tuning could be in different cortical layers. Our analyses uncovered unique information from multiple feedforward and feedback models, and models from each had different layer profiles. In future, we would like to separate voxel tuning into more than just the balance between sensory

and predictive signals. The balance of different types of information in cortical layers might provide further clues to the underlying functional circuitry. Additionally, the analyses in this chapter can be extended by including more diverse computational and categorical models. For instance, our results from Chapter 3 showed that the mid-level visual H-MAX model described responses in portions of Occluded V1. Due to computational limitations related to computing H-MAX features, this model was not included in the battery of models in the current chapter, but future analyses should strive to include a wide array of models varying in computational complexity to uncover tuning properties of feedforward and feedback signals.

In this chapter, we have expanded on results from previous chapters to show that cortical depth-dependent V1 responses exhibit predictive and high-level response properties which are most prevalent in superficial layers of cortex, where many feedback connections terminate in V1. Importantly, predictive and high-level responses are unique in their information content from feedforward orientation and spatial frequency properties typically associated with V1 responses. Overall, these findings suggest that feedback connections terminating in superficial layers provide V1 neurons with contextual and associative information not available via localized feedforward input.

4.4 Methods

4.4.1 Participants

Three healthy individuals (ages 26, 23, 24) with corrected-to-normal vision gave written informed consent to participate in this study, in accordance with the institutional guidelines of the local ethics committee of the College of Science & Engineering at the University of Glasgow (#CSE300160103).

4.4.2 Stimuli

384 real-world scenes were chosen from the SUN database (Xiao et al., 2010). In order to maximize power in encoding models, scenes were chosen to be maximally uncorrelated in their spatial structure. We randomly generated 5,000 sets of 96 scenes, decomposing them using a Gabor wavelet pyramid with 8 orientations, 6 spatial frequencies, and 2 phases. Gabor filters were arranged on a Cartesian grid and spaced by double the bandwidth of the filter at each frequency. Thus, higher frequency filters were closer together than low frequency filters. Decomposed images within each set were then correlated, and the 3 sets with the lowest average Fisher Z-transformed Pearson correlations were chosen as stimulus sets. One set was presented with the lower-right quadrant occluded by a white box, one set was presented without occlusion, and one set was presented with and without occlusion.

Images were displayed in grayscale, matched for global luminance, and masked with a circular aperture which linearly faded to the background (mean grayscale across scenes)

from 4.9° to 5.15° visual angle (Kay et al., 2008). Stimuli were presented on a rear-projection screen using a projector system (768 × 768 resolution, 60 Hz refresh rate). Stimuli spanned 10.38 × 10.38° visual angle. A centralized fixation checkerboard (9 x 9 pixels) marked the center of the scene images.

4.4.3 Experimental Design

Each of the 48 event-related experimental runs consisted of 16 blocks of 3 scene presentations. Each image was flashed at 2.5Hz for 1 second, followed by 5 seconds baseline. After each block of 3 scene presentations, a null trial consisting of 6 seconds baseline was presented. Each scene was presented 3 times in two runs, totaling 6 presentations over the experiment. Scene presentation orders were pseudo-randomized so that each scene appeared as the first trial in a block once, no scene could appear in two consecutive blocks, and the temporal design was optimized based on Dale 1999. To ensure fixation, we instructed participants to respond via a button press to a temporally random fixation color change. So that participants would attend to the scenes, participants were asked to report whether scenes were man-made or natural scenes. We conducted 5 runs of retinotopic mapping (polar-angle, eccentricity, and 3 runs of moving bars in 8 directions) separately from the main experiment.

4.4.4 fMRI Acquisition

MRI data were collected at the Scannexus, University of Maastricht, Maastricht, Netherlands over five scanning sessions using a research-dedicated 7T Magnetom MRI system (Siemens, Erlangen, Germany) with a 32-channel head coil (Nova Medical Inc.; Wilmington, MA, USA). Five T1-weighted anatomical datasets were acquired using a 3D-MPRAGE sequence (256 sagittal slices, matrix = 384 × 384, voxel size = 0.6mm isotropic). To correct for inhomogeneities, we collected a gradient echo proton density (GE-PD) scan with the same parameters as the MPRAGE acquisition during each session.

High-resolution functional images were obtained using a T2*-weighted gradient echo EPI with the following parameters: echo time (TE) = 25ms, repetition time (TR) = 2000ms, iPAT-factor = 3, multi-band factor = 2, flip angle = 75°, number of slices = 56, matrix = 186 × 186, voxel size = 0.8mm isotropic. The field-of-view included occipital early visual cortex, centered on the Calcarine Sulcus. To correct for EPI distortions additional functional volumes (five volumes in the encoding direction and five volumes with a reversed encoding direction) were acquired before each experimental run.

4.4.5 MRI Data Preprocessing

Functional and anatomical images were analyzed using BrainVoyager QX (version 2.8; Brain Innovation; Maastricht, Netherlands) and custom code in MATLAB (version 2016b; The MATHWORKS Inc.; Natick, MA, USA). Anatomical images were corrected for bias field

inhomogeneities by dividing the Proton Density images by the T1-weighted images and interpolated to a nominal voxel size of 0.4mm isotropic so that functional data would have a voxel geometry twice that of anatomical images. Gray matter/white matter and gray matter/CSF boundaries were detected using a combination of automatic and manual segmentation tools from BrainVoyager QX, ITK-SNAP (Yushkevich et al., 2016), and Segmentator software (Gulban and Schneider, 2017). These boundaries were used to define an inflatable cortical mesh and define 6 cortical depth ROIs within V1 (Depths = 90%, 74%, 58%, 42%, 26% and 10% gray matter depth; Brainvoyager QX).

To correct distortions in functional images we recorded 5 functional volumes with reversed phase encoding compared to experimental runs. In these pairs of images distortions go in opposite directions and we used them to estimate the susceptibility-induced off-resonance field using a method described in Andersson et al., 2003, as implemented in the FSL software package (Smith et al., 2004). After aligning all functional runs to the first run using 3D rigid body motion correction, the estimated off-resonance field was used to correct for EPI distortions. Functional runs were co-registered to the individual anatomical scan with a rigid body (6 parameter) transformation. No spatial or temporal smoothing was performed. Retinotopic data were overlaid onto their respective anatomical data in the form of an inflated surface and were used to define early visual areas V1 using linear cross-correlation of eight polar angle conditions based on Polar Angle mapping.

4.4.6 Population Receptive Field Mapping

We performed Population Receptive Field (pRF) modeling on all V1 voxels (Dumoulin and Wandell, 2008). We defined potential pRF models as two-dimensional isotropic Gaussian functions with 24 logarithmically-spaced sizes (σ) between 0.1° visual angle and the size of the stimulus (5.19°). They had spatial positions based on a Cartesian grid where spacing between potential models was 1σ for each size. Additionally, positions were only within the size of the stimulus (i.e. $< 5.19°$ visual angle). Potential model time courses were calculated by taking the dot product of each model with a binarized stimulus time course and convolving with a canonical 2γ hemodynamic response function.

In addition to the predicted model time course, models included polynomial confound predictors (degree 0 through half the number of minutes of each functional run). Each model was fit to 3 of 5 retinotopic mapping runs and tested on the left-out two. Data were folded over every combination of training and testing runs. Model fitting used Ordinary Least Squares with Bootstrap Aggregation (Breiman, 1996); 50 random samples of 60% of the training data were used to derive model parameters, which were then averaged. The use of Bootstrap Aggregation has two advantages: it acts as a regularizer on the model estimation by averaging over unstable parameters, and it decreases the autoregressive components of the model. This method has shown to outperform autoregressive time series models in economic data (Inoue and Kilian, 2004). Each voxel's pRF was defined as the model with the

highest average R^2 value (with confound variables held constant).

4.4.7 Encoding models

In each voxel, we defined voxel-specific predictors relating to feedforward and feedback information streams. Feedforward models were defined as model responses to scenes as they were *presented*. In other words, if a V1 neuron's receptive field was located in the lower-right quadrant of the scenes, it would receive meaningful sensory input during Non-Occluded scene presentations, but would not receive input during Occluded scenes. For feedforward models, we calculated the responses of two low-level computational models (Weibull and Gist; see individual sections for detail) based on the portion of each scene spanned by the voxel's pRF exactly as scenes were presented (see Figure 4.1).

Feedback models were defined as model responses as they would be *predicted*. These models treated each scene as if it was Non-Occluded whether it was presented Occluded or not. We therefore calculated the pRF-specific responses of the Weibull and Gist models based on Non-Occluded versions of each scene. A high-level Category model was also included as a feedback model. In total, each model consisted of 87 voxel-specific predictors (2 FF Weibull, 2 FB Weibull, 32 FF Gist, 32 FB Gist, and 19 Category predictors). Voxels with pRFs in the Non-Occluded portions of scenes did not include FB Weibull or FB Gist models because predictors were identical in feedforward and feedback cases due to lack of occlusion.

Models also included polynomial confound predictors (degree 0 through half the number of minutes of each functional run). Models were fitted using Ordinary Lease Squares regression to voxel time courses of 40 runs (320 scenes) and tested on the remaining 8 runs of data (64 scenes). Bootstrap Aggregation was used with 50 random samples of 50% of time points to reduce autoregressive model components and to increase estimated parameter generalizability (Breiman, 1996; Inoue and Kilian, 2004). This process was repeated over 1000 cross-validation folds. In each fold of the data, we calculated semi-partial correlation statistics for each individual model (Equation 3.4.3) to determine the unique correlation of each model.

4.4.8 Weibull Model

The Weibull image contrast model measures the distribution of contrast values for an image and seeks to emulate the X and Y cells in the Lateral Geniculate. It has a two-dimensional output. We used the Weibull image contrast model outlined in Scholte et al. (2009) and Groen et al. (2012, 2013). Par and Mag components of the model (X and Y cell models, respectively) were calculated as maps. Voxelwise contrast energy (the first component of the model) was estimated as the weighted mean of Par maps based on each voxel's pRF (Equation 3.2). Spatial coherence (the second component of the model) was estimated as the

weighted variance of the Mag maps based on each voxel's pRF (Equation 3.3).

4.4.9 Gist Model

The Gist algorithm (Oliva and Torralba, 2001) measures the distribution of oriented bandpass Gabor filter responses in localized portions of images. Our model consisted of 8 orientations, and 4 spatial frequencies. We calculated maps for each feature, resulting in 32 maps. Voxelwise Gist features were then calculated as the weighted mean of the maps based on each voxel's pRF (Equation 3.2).

4.4.10 Category model

In addition to the lower-level Weibull and Gist models, we included a high-level category model. This model was based on the hierarchical organization of the SUN database (Xiao et al., 2010). We included predictors for the two highest levels of the hierarchy (19 total predictors):

- Indoor:
  - Shopping and Dining
  - Workplace (office building, factory, lab, etc.)
  - Home or hotel
  - Transportation (vehicle interiors, stations, etc.)
  - Sports and leisure
  - Cultural (art, education, religion, military, law, politics, etc.)
- Outdoor Natural
  - Water, ice, and snow
  - Mountains, hills, desert, and sky
  - Forest, field, and jungle
  - Man-made elements
- Outdoor Man-made
  - Transportation (roads, parking, bridges, boats, airports, etc.)
  - Cultural or historical building/place (military, religious)
  - Sports fields, parks, and leisure spaces
  - Industrial and construction
  - Houses, cabins, gardens, and farms
  - Commercial buildings, shops, markets, cities, and towns

4.4.11 Retinotopic projections

Using each voxel's two-dimensional Gaussian response, as estimated in our pRF analysis, we projected voxels with significant voxelwise model encoding into visual space. Voxel pRF functions for all voxels with significant encoding were summed within each the subject's data. This procedure was repeated for each model. Feedback Weibull and Gist models (Rows 3 & 4 of Figure 4.3) only have Occluded components because these models do not exist for Non-Occluded scene presentations.

4.4.12 Voxel tuning

Tuning was calculated by averaging semi-partial correlation values for each model across validation folds on a voxelwise basis. Within each voxel, model values were divided by the sum of all model semi-partial correlations, resulting in a compositional structure (Aitchison, 1982). Since compositional analyses cannot contain negative values, any negative correlation values were assigned a value of zero. We then summed the two feedforward models to get a measure of the total tuning toward sensory input. The same was done with feedback models to get tuning toward predictive signals. We then subtracted the feedforward value from the feedback values to get a measure of each voxel's overall tuning toward sensory or predictive signals. These values ranged from -1 (100% tuning for Feedforward signals, 0% tuning for Feedback) to 1 (0% tuning for Feedforward signals, 100% tuning for Feedback), with a value of 0 indicated a 50/50 split.

To visualize the tuning of voxels within each cortical depth, we calculated histograms of feedforward-feedback tuning values within each depth with 100 bins. Probability density functions of tuning values were then defined by kernel density estimation of each histogram with a Gaussian kernel (Hill, 1985; Silverman, 1986).

Chapter 5

General Discussion

The work presented in this thesis describes neuroscientific results on the information content of cortical feedback to early visual cortex, particularly primary visual cortex. We utilized the sensitivity of functional magnetic resonance imaging (fMRI) to dendritic energy-consumption (Logothetis, 2007, 2008) to detect and characterize feedback signals in combination with an occlusion paradigm derived from that of Smith and Muckli (2010) and Muckli et al. (2015). During normal vision, both feedforward and feedback signals are present, thus a useful approach to study feedback is to isolate it from feedforward input. We occluded one quadrant of the visual field during stimulus presentation in order to remove meaningful feedforward input about scenes in a portion of retinotopic visual cortex. We used fMRI to assess brain activity in early visual cortex, allowing us to detect dendritic signaling associated with cortical feedback due to its sensitivity to cortical energy consumption .

In Chapter 2, we investigated potential high-level information in cortical feedback to V1 and V2. We presented subjects with an expanded version of the occlusion paradigm from Smith and Muckli (2010) and Muckli et al. (2015). We included twenty-four partially occluded scenes from six categories and spatial depths. These two high-level scene characteristics were chosen because they have previously been shown to modulate early visual cortical responses (Walther et al., 2009; Kravitz et al., 2011). We were therefore interested in whether these characteristics also modulate feedback to V1 and V2. We found that response patterns in these subregions contain high-level category information, but we did not find that visual depth information generalized across exemplars. Additionally, we found that retinotopic responses in Occluded V1 and V2 differed from each other, suggesting that feedback to these two areas has different information content. These results match the known anatomical connections from mid- and higher-level visual areas in the ventral stream (Rockland et al., 1994; Rockland and Ojima, 2003).

In Chapter 3, we probed the information content of Occluded V1 and V2 responses at multiple levels of complexity using Representational Similarity Analysis (RSA) and encoding models. By analyzing data from Chapter 2 in these frameworks, we were able to compare both local (voxelwise) and distributed (multi-voxel) Occluded responses to three

biologically-inspired computational models (the contrast energy-based Weibull model, the orientation-based Gist model, and the mid-level vision H-Max model), and the high-level scene characteristics explored in Chapter 2. Using RSA, we also compared scene representations from Occluded and Non-Occluded areas. We found that in Non-Occluded areas, V1 and V2 represent scenes similarly, while Occluded V1 and V2 do not. We also found that scene representations in Occluded V1 and V2 were correlated with high-level Category and H-Max models. Individual voxel encoding models showed that Occluded V1 voxels within 5° visual angle of fixation encode low-level information about the occluded scene, while voxels outside of 5° encode higher-level information. These results highlight a potential visual field bias in the type of information transmitted to V1 through feedback, with foveal voxels receiving more precise, low-level scene information, and peripheral voxels receiving more invariant or global scene features.

In Chapter 4, we examined the laminar profile of Occluded V1 using high-resolution ($0.8mm^3$) 7T fMRI. We again expanded our stimulus set, presenting 192 Occluded scenes and 192 Non-Occluded scenes. This large stimulus set allowed us to map scene information onto voxel responses in greater detail, and the use of both Occluded and Non-Occluded scenes allowed us to compare voxel responses when receiving only feedback with responses when receiving feedforward, lateral and feedback information. We found that V1 responses exhibit predictive and high-level response properties in addition to feedforward orientation and spatial frequency properties typically associated with V1 responses. These predictive and high-level responses were primarily associated with superficial layers of cortex. We also found that voxel tuning toward feedforward and feedback signals was different between cortical layers of V1. Our findings suggest that feedback connections terminating in superficial layers provide V1 neurons with contextual and associative information not available via localized feedforward input.

The neuroscientific results presented in this thesis extend our knowledge about the information content of cortical feedback to early visual cortex. These results add support to the notion that V1 can be considered to speak two languages (Muckli and Petro, 2013). Not only does it play a role as an early stage of processing of sensory visual input, where it deals with processing low-level features, but it also receives messages from diverse areas of cortex. These messages might supplement local processing by providing contextual information.

5.1 Decomposing the explanatory aspects of models

As part of our analyses of encoding models, we used semi-partial correlations to measure unique information in individual model responses that related to cortical responses from fMRI signals. Statistically, this measure is the association between voxel response and a model of interest after removal of any model variability which exists in other models, too. Measuring the unique relationships between individual models and voxel responses allowed

us to ascertain which models improved predictions and which ones did not. In the case that individual models did improve predictions, we were also able to quantify by how much. This method has distinct advantages over evaluating models individually, as responses from different models can be highly correlated with each other, and thus share explained variance in their predictions of neuronal response data. For example, imagine comparing the efficacy of three different models. If all three models perform identically in explaining a dataset, then one might assume that each is an equally good model. This is in fact true, but does not always tell the complete story. Suppose semi-partial correlations are then computed for each model; Model *A* has the same unique explanatory power as it did when tested individually, while Models *B* and *C* have no unique explanatory power. We now know that Model *A* explained a portion of the data completely unique from that of Models *B* and *C*, and we know that the portion of the data explained by Models *B* and *C* was identical. These three models are still equally good at explaining the data, but we now know that Models *B* and *C* are redundant in doing so.

In this extreme example, if Models *A*, *B* and *C* were two feedforward and one feedback model, respectively, then there is no evidence for feedback playing a role in processing. Feedforward models already explain everything that can be explained in the data. If, however, Model *A* was a feedback model, then something has been added by its inclusion. This was the case in Chapters 3 and 4 of this thesis. We found that extra explanatory power was afforded by including models of predictive and contextual feedback that could not be explained by feedforward models alone. However, the example we have provided is an intentional oversimplification. It is relatively rare for any models to be completely uncorrelated with other models unless they are designed that way. Therefore in future analyses, it will be important to decompose the explanatory power of models to understand better which aspects are shared and which are unique. This has previously been accomplished with low numbers of models (Lescroart et al., 2015; Ince et al., 2016; Giordano et al., 2017), but this process becomes increasingly difficult as the number of models under consideration grows. Despite the challenges associated with variance partitioning, understanding the amount of explanatory power that is shared between models could help researchers develop more effective models of sensory processes in cortex.

5.2 Deep learning models for studying feedback to early visual cortex

One promising class of models for studying the brain is that of deep learning. Deep learning models are from the field of artificial intelligence and are loosely based on cortical architectures (LeCun et al., 2015). By mimicking aspects of brain function, deep learning can rival, and even outperform, human beings on a number of limited tasks, such as image recognition (Krizhevsky et al., 2017), motor control (Mnih et al., 2015), and speech recognition (Hinton

et al., 2012). Moreover, deep networks can better explain cortical neuronal recordings in humans or non-human primates than existing neuroscientific models (Khaligh-Razavi and Kriegeskorte, 2014; Cadieu et al., 2014). Despite some biologically-unrealistic aspects of modern deep learning implementations, these examples strongly suggest that deep learning captures something important about how the cortex works (Marblestone et al., 2016).

With this sentiment in mind, it is logical that deep learning algorithms could be used as a general framework to devise computational hypotheses which can be tested against experimental data. Like the sensory systems of the brain, deep learning models stack multiple layers of non-linear computations on top of each other in order to learn complex statistical dependencies in input data. As previously mentioned, this has been implemented with great success to explain aspects of higher-order brain function (Serre et al., 2007; Khaligh-Razavi and Kriegeskorte, 2014; Cadieu et al., 2014; Güçlü and van Gerven, 2014). More relevant to this thesis, however, is a growing body of work dedicated to the understanding the general role of feedback in cortical circuitry using network models (see Reichert and Serre, 2013; Spoerer et al., 2017). One interesting line of inquiry is the development of multi-compartment neuron models, which implement segregated dendrites for more complicated computations than simple integrate-and-fire neural models are capable of (Klijn et al., 2017). Recently, Naud and Sprekeler (2017) fitted multi-compartment biophysical models of layer 5 pyramidal neurons to patch-clamp data collected by Larkum et al. (1999). This work showed that basal and apical dendritic inputs to pyramidal neurons are both recoverable from the output of these cells, meaning that neurons might use signal multiplexing to deliver both their feedforward and feedback information to other neurons. If true, this exciting work would add credence to the applicability of deep learning models to the study of brain function, as it shows that cortical neurons have a built-in capability for error propagation (the principle that endows deep learning its ability to learn through backpropagation; LeCun et al., 2015; Guergiuev et al., 2016; Lillicrap et al., 2016).

In Chapter 3 of this thesis, our comparisons of models to brain data included multiple levels of model complexity. Due to the hierarchical nature of deep learning networks, these models are ideal candidates for naturally spanning multiple levels of complexity, and thus could be used similarly. By calculating the unique information from each layer which can be used to describe voxel responses, we might be able to learn more about the retinotopic organization of feedback to early visual cortex. We found that feedback signals sent to foveal portions of V1 share information with low-level properties of occluded scenes, while feedback to peripheral portions of V1 are more associated with high-level scene properties. This finding suggests that the pattern of connections to V1 is retinotopically specific, but raises a question of how topological this specificity might be. It is possible that there is a gradient of complexity in feedback to V1, with low-level information to the fovea smoothly giving way to more and more complex information as eccentricity increases. Alternatively, the topology might be more segregated, with low-level information being sent homogeneously to foveal areas up to a particular eccentricity, with a sharp switch to higher-level information outside

of this region.

We have described how diverse levels of complexity in deep learning networks could help to clarify information encoding of feedback signals, but these arguments equally hold for feedforward information encoding. In Chapter 3, Non-Occluded areas encoded information from all models, but were dominated by low-level Weibull and Gist models, while voxels in Occluded areas preferred higher-level information. In Chapter 4, where feedforward and feedback models are separable based on voxels seeing scenes with and without occlusion, we could test whether the same holds true. It would be interesting to examine the balance of information complexity in different cortical layers of V1. Deep learning models provide a means of testing these and many other hypotheses about the organization of retinotopic cortex.

5.3 Studying cortical layer function in humans

Ultra high-field, high-resolution fMRI can measure functional activity in human cortex at the level of layers (Olman et al., 2012) and columns (Yacoub et al., 2008), providing unprecedented views of cognitive processes in the human brain. Layer-specific imaging at ultra-high field is currently helping to bridge the gap between the microscopic resolution of neurophysiology and large-scale sampling of the active brain in functional imaging (Dumoulin et al., 2017). Furthermore, layer-specific fMRI is helping to draw inference about the direction of information flow in the brain. In Chapter 4 of this thesis, we utilized the high-resolution capabilities of 7T fMRI to measure feedback signals sent to specific layers of V1. We found that predictive and contextual signals are sent to superficial layers of early visual cortex, building on previous layer-specific results of occlusion (Muckli et al., 2015).

While there has been substantial technological maturity of ultra high-field fMRI in recent years, effective use of high-resolution fMRI is still highly specialized (Kashyap et al., 2017; Uludağ and Blinder, 2017). Image acquisition is more complicated at magnet strengths greater than or equal to 7 Tesla due to complications regarding static and radio-frequency field inhomogeneities (Vaughan et al., 2001), which can cause image artifacts and distortions. The so-called Nyquist ghosting effect in functional images using EPI readouts can also affect image quality at high-field (Yarach et al., 2017). Additionally, higher spatial resolution increases the possibility that participant motion will affect anatomical image quality (Trampel et al., 2017) from which cortical depths are defined. Analysis choices can also greatly impact the quality of cortical depth-based results (Kashyap et al., 2017). One example cause of this is that high-resolution anatomical and functional images often do not match because of image distortions due to differing image readouts.

The increased overhead required to collect high-quality 7T data and analyze it effectively calls for pragmatic study designs. For instance, data collected in Chapter 4 of this thesis was from a single subject who was scanned in five sessions. This design allowed us to examine

the responses of different cortical layers of one brain area in great detail (V1). Since we were able to collect anatomical data in each session, we had substantial information from multiple anatomical image contrasts to differentiate different tissue classes. This plays an important role in our ability to manually identify vasculature and separate it from the tissue of interest, gray matter. As neuroscientific studies continue to move toward examining brain mechanisms at the mesoscopic scale (Martino et al., 2017), it will be increasingly important to focus efforts on understanding the characteristics of the individual brain being studied.

An important factor in future studies investigating cortical layers using fMRI will be the magnetic properties of the underlying tissues. Many groups have shown that different layers are differentiable based on their tissue properties (Trampel et al., 2017; Lifshits et al., 2017; Yen et al., 2017). These tissue properties, spin-lattice relaxation time (T_1), spin-spin relaxation time (T_2^*), and proton density (or baseline magnetism; S_0), are the properties that determine signal in fMRI. It will therefore be crucial to map these properties along with vasculature considerations as experiments achieve higher resolutions through cortical layers. Yu et al. (2013) showed that it is possible to achieve 50μm spatial resolution (and 50ms temporal resolution) through cortical layers in functional imaging of mice by focusing imaging on only a small column and turning off any phase encoding steps in acquisition. With such high-resolution functional scanning, it would be important to tune sequences toward the tissue properties of layers of interest. Additionally, Tian et al. (2010) showed by mapping layer-specific microvasculature using two-photon microscopy in mice and then recording activity using fMRI that cortical layers have predictable onset latency differences. This shows the importance of understanding vasculature in human cortex for layer-specific experiments.

Despite challenges set forth by technological and biological constraints, advances in tools from the fields of functional MRI, information processing and computer science are making it possible to study cortical feedback processes in greater and greater detail. This thesis has aimed to progress our understanding of some of the types of information that might be present in cortical feedback to early visual cortex by leveraging these tools. We have shown that it is possible to isolate and map information onto responses in retinotopic cortex not receiving meaningful feedforward input, and that there is much to be learned from studying responses associated with cortical feedback.

Appendix A

Supplemental Data for Chapter 2

	V1			V2		
	Ind. Scenes	Category	Depth	Ind. Scenes	Category	Depth
Upper-right quadrant (Non-Occluded)	54.72 [48.8 60.8]	40.3 [36.9 43.9]	22.2 [20.6 24.1]	45.6 [39.3 52.3]	35.0 [31.2 28.4]	20.5 [19.2 21.8]
Lower-left quadrant (Non-Occluded)	51.3 [44.6 57.3]	36.6 [32.1 40.6]	17.7 [15.6 19.5]	40.1 [33.5 45.7]	33.6 [30.4 37.1]	13.9 [11.5 15.8]
Lower-right quadrant (Occluded)	6.4 [4.9 8.1]	8.5 [6.5 10.3]	5.32 [3.7 6.9]	3.8 [2.8 4.8]	6.7 [5.5 8.0]	3.7 [2.4 5.1]

Table A.1: Quantitative results from SVM classification analyses. Related to Figure 2.2. Classification results, reported as percent above chance-level (chance-level = 4.17%, 16.67%, and 50% for individual scenes, category and depth, respectively). 1000 bootstrap samples of mean performance were drawn from individual subject results, and 95% confidence intervals are shown in brackets.

	V1		V2	
	Category	Depth	Category	Depth
Upper-right quadrant (Non-Occluded)	9.6 [8.7 10.6]	7.7 [5.3 10.0]	9.1 [8.0 10.0]	8.1 [6.1 9.8]
Lower-left quadrant (Non-Occluded)	6.3 [5.2 7.5]	-0.2 [-2.2 1.6]	8.9 [7.5 10.0]	-1.7 [-3.7 0.0]
Lower-right quadrant (Occluded)	3.0 [2.1 4.1]	0.5 [-0.9 2.0]	2.2 [1.5 2.9]	-0.5 [-2.1 1.0]

Table A.2: Quantitative results from SVM cross-classification analyses (lower-visual field occluder).

	V1		V2	
	Category	Depth	Category	Depth
Upper-right quadrant (Non-Occluded)	1.9 [1.0 3.1]	1.1 [-1.4 3.6]	1.9 [0.1 3.9]	0.9 [-0.1 2.6]
Lower-left quadrant (Non-Occluded)	9.9 [4.9 14.5]	1.0 [-4.5 6.5]	9.3 [5.3 13.1]	-3.9 [-9.4 0.9]
Lower-right quadrant (Occluded)	4.1 [2.2 6.4]	2.5 [0.6 4.9]	4.7 [2.6 6.5]	0.5 [-3.2 4.0]

Table A.3: Quantitative results from SVM cross-classification analyses (upper-visual field occluder). Related to Figure 2.2. Cross-classification results (Main experiment and Experiment 2), reported as percent above chance-level (chance-level = 16.67% for category and 50% for depth). For analysis of category, 18 (of 24) scenes were selected, leaving out one scene per category. For depth, 22 (of 24) were selected, leaving out one scene per depth. The classifier was tested on the remaining scenes in a cross-classification approach. Due to the large number of possible image permutations in these analyses, we randomly assigned scenes to training and testing sets 100 times in each subject. 1000 bootstrap samples of mean performance were drawn from individual subject results, and 95% confidence intervals are shown in brackets.

		V1			V2	
	Ind. Scenes	Category	Depth	Ind. Scenes	Category	Depth
Upper-right quadrant (Non-Occluded)	18	18	18	18	18	18
Lower-left quadrant (Non-Occluded)	18	18	18	18	18	17
Lower-right quadrant (Occluded)	17	16	10	14	16	7

Table A.4: Subjects with significant individual subject classifications (permutation). Number of subjects (from N=18) with significant individual subject classifications based on permutation testing (1000 random permutations of training data labels). In individual subjects, significance was determined by comparison to a null distribution of 1000 permutations, $p < 0.05$ considered significant.

	V1			V2		
	Ind. Scenes	Category	Depth	Ind. Scenes	Category	Depth
Upper-right quadrant (Non-Occluded)	18	18	18	18	18	18
Lower-left quadrant (Non-Occluded)	18	18	18	18	18	17
Lower-right quadrant (Occludened)	15	17	9	13	15	7

Table A.5: Subjects with significant individual subject classifications (Wilcoxon signed-rank). Number of subjects (from N=18) with significant individual subject classifications based on Wilcoxon signed-rank testing (classification performance from 8 cross-validated folds compared to chance-level).

	V1		V2	
	Category	Depth	Category	Depth
Upper-right quadrant (Non-Occluded)	18	15	18	15
Lower-left quadrant (Non-Occluded)	18	5	18	5
Lower-right quadrant (Occluded)	15	5	14	6

Table A.6: Subjects with significant individual subject cross-classification (Wilcoxon signed-rank). Number of subjects (from N=18) with significant individual subject cross-classification performances tested using Wilcoxon signed-rank.

We tested whether the number of voxels in V1 was greater than the number in V2 (Figure A.1). There was a significant difference in voxel counts for Upper-Right and Lower-Left quadrants, but not in the Lower-Right Occluded quadrant (p = 0.007, 0.016, and 0.116, respectively, one-sided paired t-tests). Occluded areas were the only quadrant to not have significant differences in voxel counts between V1 and V2. We conclude that differences between Occluded V1 and V2 are not due to differences in voxel counts. We mapped the voxel density from our visual space analyses using the inverse of Scott's Rule-of-Thumb (Scott, 2009) for the bandwidth of a 2-dimensional kernel density estimator:

$$h^* = \bar{\sigma} n^{\frac{1}{6}} \tag{A.1}$$

Where h^* is the weighted average of pRF sizes based on a Gaussian kernel with $\sigma = 1^\circ$ of visual angle, and n is the weighted number of voxels based on the same kernel. The kernel was convolved with the image and multiplied by a map of voxel coverage. Figure S1B shows the group average (N=18) of individual subject maps.

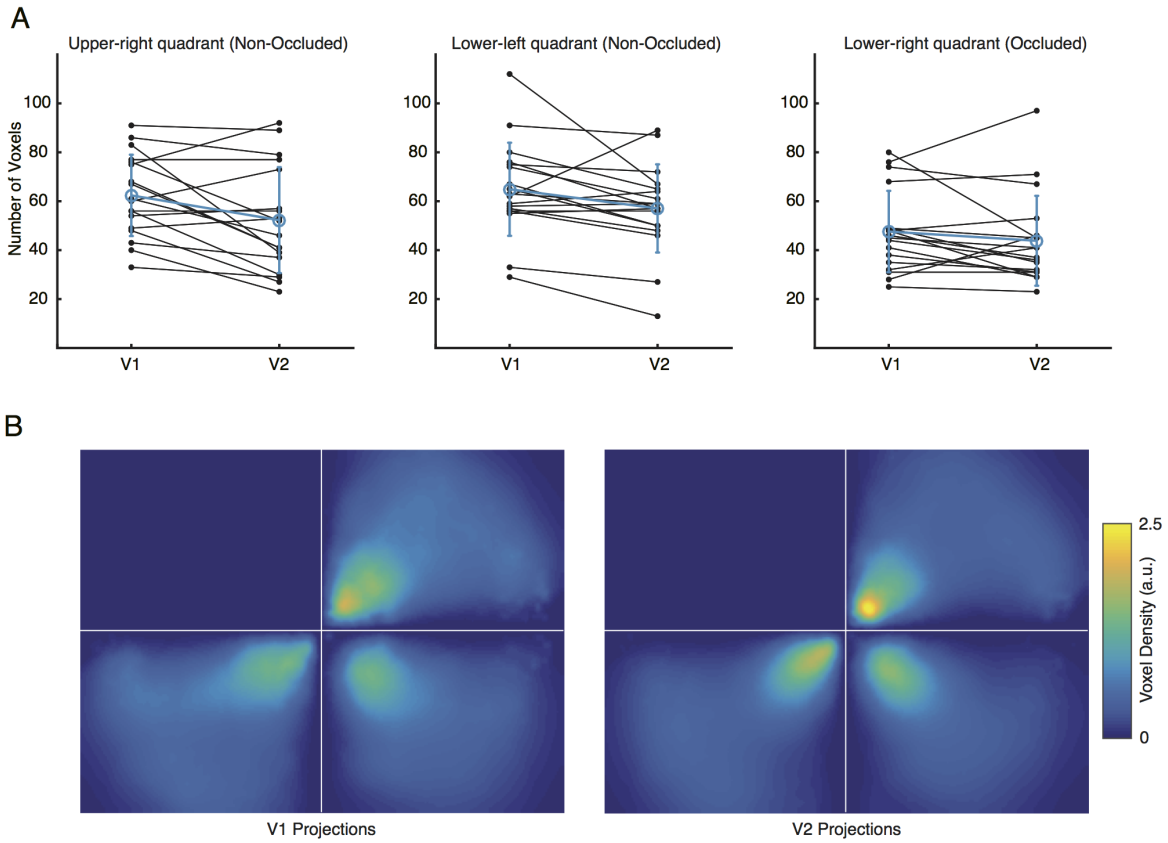

Figure A.1: Related to Figure 2.1. (A) V1 and V2 voxel counts in each region of interest. Black points and lines indicate individual subject voxel counts in V1 and V2 in each region of interest, and group statistics are shown in blue (with standard error). (B) Voxel density in regions of interest in visual space. Voxel density maps were calculated using the inverse of Scott's Rule-of-Thumb (kernel width of 1° visual angle) in individual subjects, masked by voxel coverage maps, and averaged across subjects.

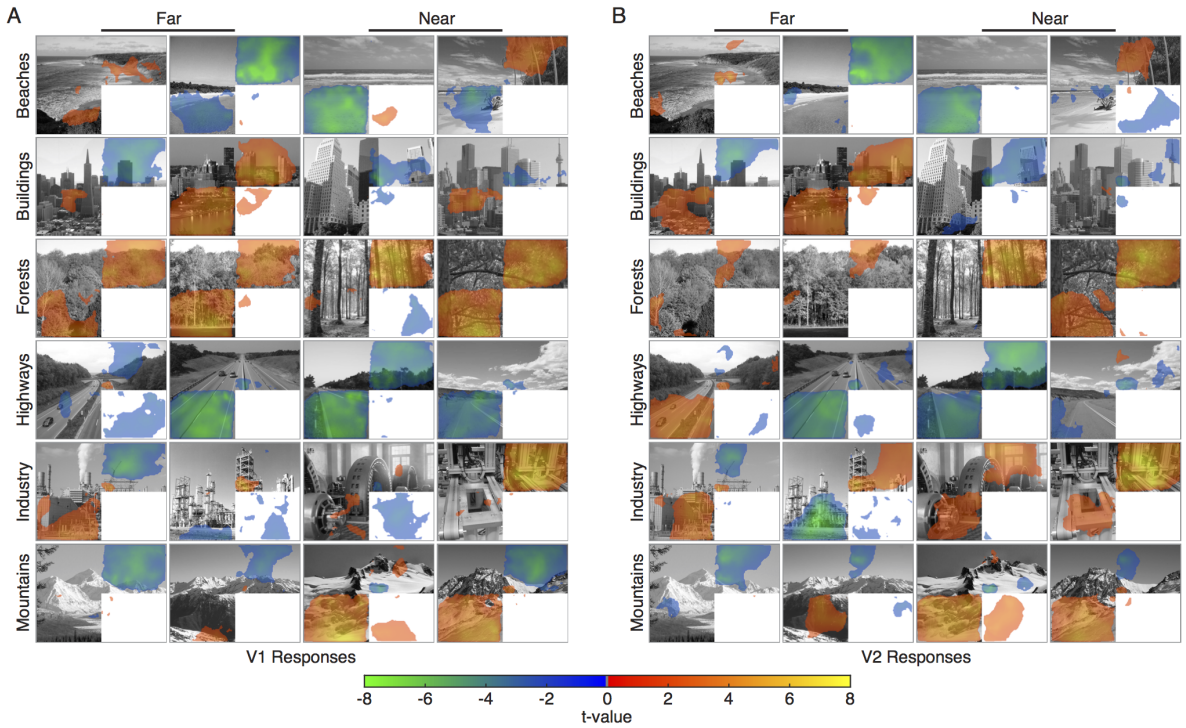

Figure A.2: Related to Figure 2.4. Projections of (A) V1 and (B) V2 response patterns into visual space. Voxel responses versus baseline were projected to visual space by calculating a weighted average of all voxels' pRFs, where weights were each voxel's response amplitude with the mean response to all scenes removed. Projections were mapped in individual subjects and a two-tailed t-test was conducted across subjects at each pixel location in visual space to obtain t-value maps ($p<0.05$ threshold). Warm colors are above mean responses, and cool colors are below.

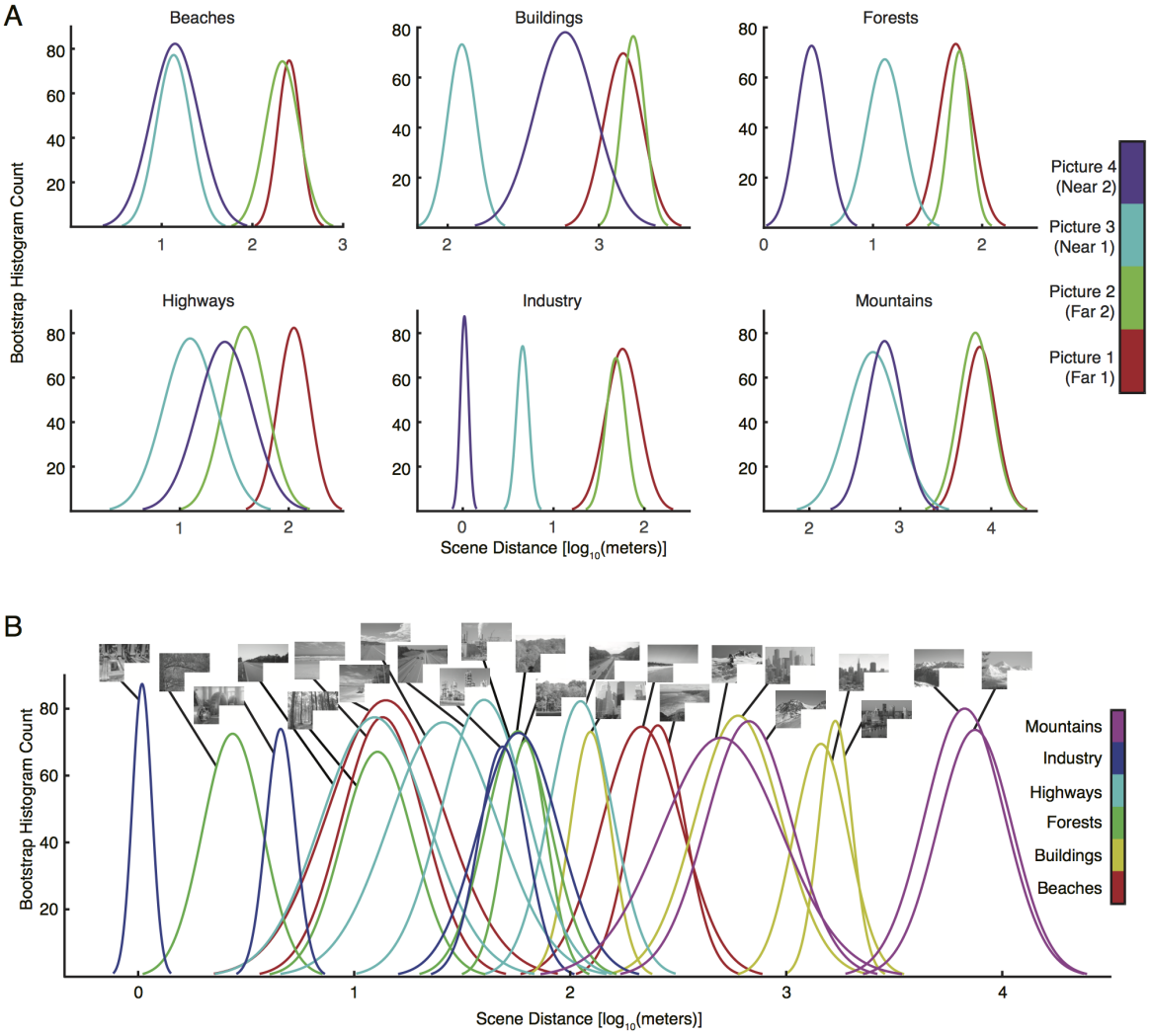

Figure A.3: Related to Figure 3.1. (A) Absolute depth ratings of scenes organized by category. Absolute depth ratings ($\log_{10}$(meters)) for our 24 scenes are shown, grouped by category. Depth ratings were obtained a separate behavioral experiment (N=10). (B) Absolute depth ratings of full scene set. Absolute depth ratings ($\log_{10}$(meters)) for our 24 scenes (scenes are identical to (A), but plotted together).

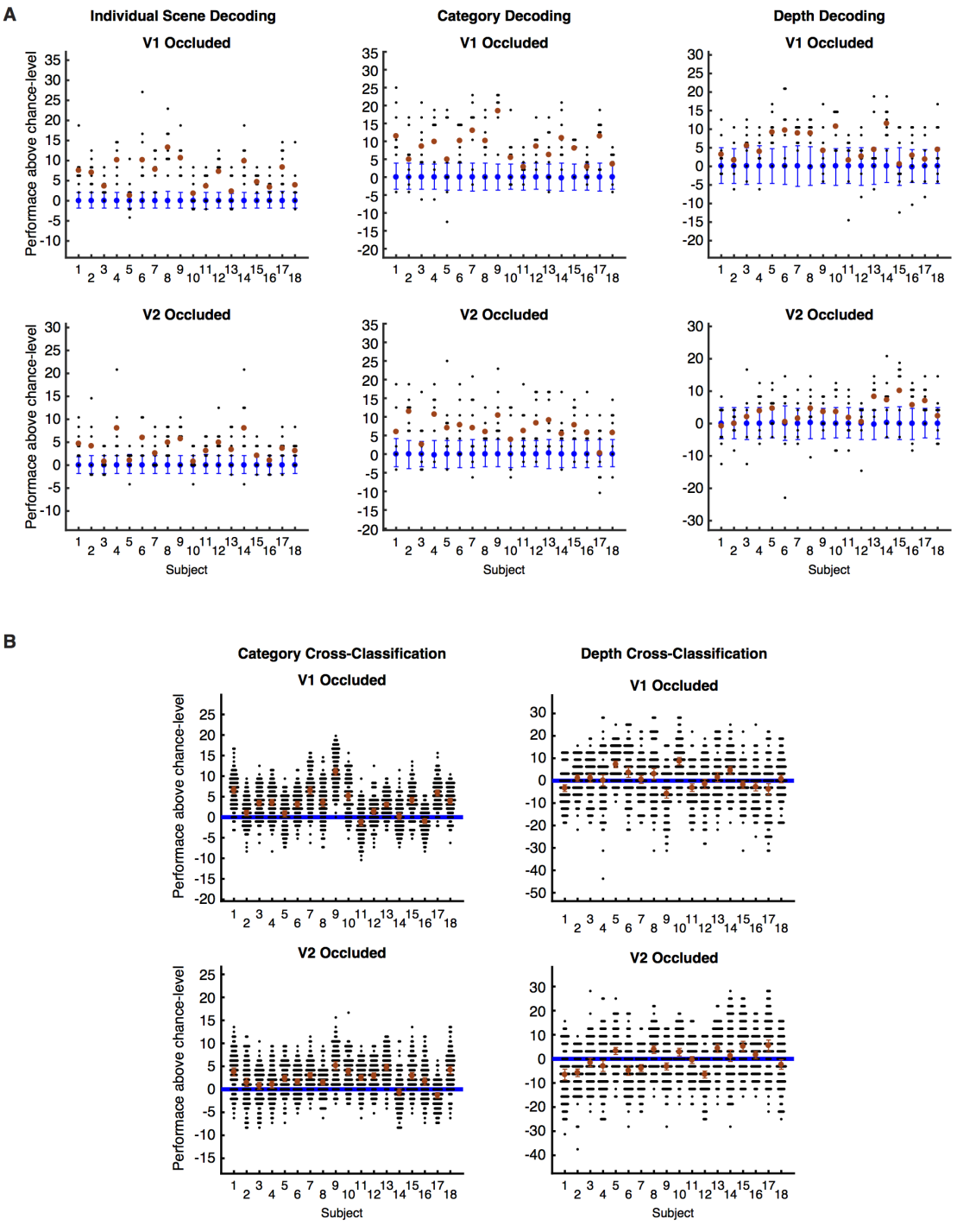

Figure A.4: Related to Figure 2.2. (A) Individual results from SVM decoding analyses. Classification performances are displayed as compared to chance-level for each analysis type. Performances from each leave-one-run-out fold are shown as black dots, with the mean of all folds shown in red. 95% confidence intervals on a permutation-based null distribution are shown in blue. (B) Individual results from SVM cross-classification analyses. Individual performances on random splits of the scene set pare shown as black dots, and 95% confidence intervals on the mean performance are shown in red. Chance-level performance is shown as a blue line.

Appendix B

Supplemental Data for Chapter 4

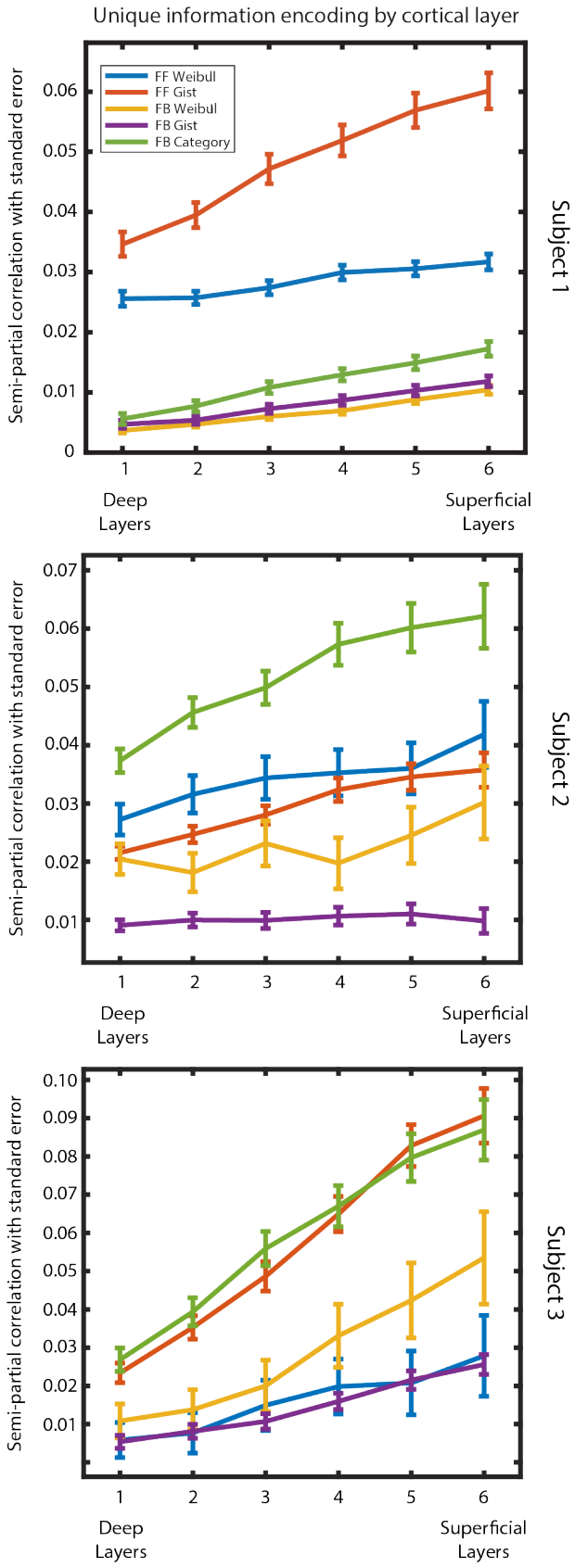

Figure B.1: Voxelwise unique model information encoding by cortical depth. Unique model information is displayed as the average of all individual voxel semi-partial correlations of model predictors while all other models are held constant. Measurements are shown for regions of interest at 6 cortical depths [10%, 26%, 42%, 58%, 74%, 90%], with Layer 1 being labeled as the deepest layer and Layer 6 as the most superficial. Lines indicate the mean correlation and error bars indicate standard error.

Bibliography

J. Aitchison. The statistical analysis of compositional data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 44(2):139–177, 1982. ISSN 00359246. URL http://www.jstor.org/stable/2345821.

A. Alink, W. Singer, and L. Muckli. Capture of auditory motion by vision is represented by an activation shift from auditory to visual motion cortex. *J Neurosci*, 28(11):2690–7, 2008. ISSN 1529-2401 (Electronic) 0270-6474 (Linking). doi: 10.1523/JNEUROSCI.2980-07.2008. URL http://www.ncbi.nlm.nih.gov/pubmed/18337398.

A. Alink, C. M. Schwiedrzik, A. Kohler, W. Singer, and L. Muckli. Stimulus predictability reduces responses in primary visual cortex. *J Neurosci*, 30(8):2960–6, 2010. ISSN 1529-2401 (Electronic) 0270-6474 (Linking)>. doi: 10.1523/JNEUROSCI.3730-10.2010. URL http://www.ncbi.nlm.nih.gov/pubmed/20181593.

J. L. Andersson, S. Skare, and J. Ashburner. How to correct susceptibility distortions in spin-echo echo-planar images: Application to diffusion tensor imaging. *NeuroImage*, 20 (2):870–888, 2003. doi: 10.1016/s1053-8119(03)00336-7. URL https://doi.org/10.1016/s1053-8119(03)00336-7.

A. Angelucci and P. C. Bressloff. Contribution of feedforward, lateral and feedback connections to the classical receptive field center and extra-classical receptive field surround of primate v1 neurons. *Progress in Brain Research*, 154(A):93–120, 2006. doi: 10.1016/S0079-6123(06)54005-1.

A. Angelucci and J. Bullier. Reaching beyond the classical receptive field of v1 neurons: Horizontal or feedback axons? *Journal of Physiology-Paris*, 97(2-3):141–154, 2003. doi: 10.1016/j.jphysparis.2003.09.001. URL https://doi.org/10.1016/j.jphysparis.2003.09.001.

A. Angelucci, J. B. Levitt, E. J. S. Walton, J.-M. Hupé, J. Bullier, and J. S. Lund. Circuits for local and global signal integration in primary visual cortex. *Journal of Neuroscience*, 22(19):8633–8646, 2002. ISSN 0270-6474. URL http://www.jneurosci.org/content/22/19/8633.

H. Ban, H. Yamamoto, T. Hanakawa, S. i. Urayama, T. Aso, H. Fukuyama, and Y. Ejima. Topographic representation of an occluded object and the effects of spatiotemporal context in human early visual areas. *Journal of Neuroscience*, 33(43):16992–17007, 2013. doi: 10.1523/jneurosci.1455-12.2013. URL https://doi.org/10.1523/jneurosci.1455-12.2013.

H. B. Barlow. Action potentials from the frog's retina. *The Journal of Physiology*, 119(1): 58–68, 1953a. doi: 10.1113/jphysiol.1953.sp004828. URL https://doi.org/10.1113/jphysiol.1953.sp004828.

H. B. Barlow. Summation and inhibition in the frog's retina. *The Journal of Physiology*, 119 (1):69–88, 1953b. doi: 10.1113/jphysiol.1953.sp004829. URL https://doi.org/10.1113/jphysiol.1953.sp004829.

H. B. Barlow. *Possible Principles Underlying the Transformations of Sensory Messages*, pages 216–234. Sensory Communication. The MIT Press, 1961a. doi: 10.7551/mitpress/9780262518420.003.0013. URL https://doi.org/10.7551/mitpress/9780262518420.003.0013.

H. B. Barlow. *Three points about lateral inhibition*, pages 782–790. Sensory Communication. The MIT Press, 1961b.

A. M. Bastos, W. M. Usrey, R. A. Adams, G. R. Mangun, P. Fries, and K. J. Friston. Canonical microcircuits for predictive coding. *Neuron*, 76(4):695–711, 2012. ISSN 0896-6273. doi: DOI10.1016/j.neuron.2012.10.038. URL <GotoISI>://WOS:000311977900005.

C. M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006. ISBN 0387310738.

E. Borra and K. S. Rockland. Projections to early visual areas v1 and v2 in the calcarine fissure from parietal association areas in the macaque. *Front Neuroanat*, 5:35, 2011. ISSN 1662-5129 (Electronic) 1662-5129 (Linking). doi: 10.3389/fnana.2011.00035. URL https://www.ncbi.nlm.nih.gov/pubmed/21734867.

L. Breiman. Bagging predictors. *Machine Learning*, 24(2):123–140, Aug 1996. ISSN 1573-0565. doi: 10.1007/BF00058655. URL https://doi.org/10.1007/BF00058655.

A. A. Brewer, J. Liu, A. R. Wade, and B. A. Wandell. Visual field maps and stimulus selectivity in human ventral occipital cortex. *Nature Neuroscience*, 8(8):1102–1109, 2005. doi: 10.1038/nn1507. URL https://doi.org/10.1038/nn1507.

J. M. Budd. Extrastriate feedback to primary visual cortex in primates: a quantitative analysis of connectivity. *Proc Biol Sci*, 265(1400):1037–44, 1998. ISSN 0962-8452 (Print) 0962-8452 (Linking). doi: 10.1098/rspb.1998.0396. URL http://www.ncbi.nlm.nih.gov/pubmed/9675911.

C. F. Cadieu, H. Hong, D. L. K. Yamins, N. Pinto, D. Ardila, E. A. Solomon, N. J. Majaj, and J. J. DiCarlo. Deep neural networks rival the representation of primate it cortex for core visual object recognition. *PLoS Computational Biology*, 10(12):e1003963, 2014. doi: 10.1371/journal.pcbi.1003963. URL https://doi.org/10.1371/journal.pcbi.1003963.

M. Carandini. Do we know what the early visual system does? *Journal of Neuroscience*, 25 (46):10577–10597, 2005. doi: 10.1523/jneurosci.3726-05.2005. URL https://doi.org/10.1523/jneurosci.3726-05.2005.

A. Clark. Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(03):181–204, 2013. doi: 10.1017/s0140525x12000477. URL https://doi.org/10.1017/s0140525x12000477.

A. M. Dale. Optimal experimental design for event-related fmri. *Human Brain Mapping*, 8 (23):109–114, 1999. doi: 10.1002.

S. V. David, W. E. Vinje, and J. L. Gallant. Natural stimulus statistics alter the receptive field structure of v1 neurons. *Journal of Neuroscience*, 24(31):6991–7006, 2004. ISSN 0270-6474. doi: Doi10.1523/Jneurosci.1422-04.2004. URL <GotoISI>://WOS:000223102500017.

S. de Jong. Simpls: An alternative approach to partial least squares regression. *Chemometrics and Intelligent Laboratory Systems*, 18(3):251–263, 1993. doi: 10.1016/0169-7439(93)85002-x. URL https://doi.org/10.1016/0169-7439(93)85002-x.

W. Denk, J. Strickler, and W. Webb. Two-photon laser scanning fluorescence microscopy. *Science*, 248(4951):73–76, 1990. doi: 10.1126/science.2321027. URL https://doi.org/10.1126/science.2321027.

R. J. Douglas and K. A. Martin. A functional microcircuit for cat visual cortex. *The Journal of Physiology*, 440(1):735–769, 1991. doi: 10.1113/jphysiol.1991.sp018733. URL https://doi.org/10.1113/jphysiol.1991.sp018733.

R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification (2Nd Edition)*. Wiley-Interscience, 2000. ISBN 0471056693.

S. O. Dumoulin and B. A. Wandell. Population receptive field estimates in human visual cortex. *Neuroimage*, 39(2):647–60, 2008. ISSN 1053-8119 (Print) 1053-8119 (Linking). doi: 10.1016/j.neuroimage.2007.09.034. URL http://www.ncbi.nlm.nih.gov/pubmed/17977024.

S. O. Dumoulin, A. Fracasso, W. van der Zwaag, J. C. Siero, and N. Petridou. Ultra-high field mri: Advancing systems neuroscience towards mesoscopic human brain function. *NeuroImage*, nil(nil):nil, 2017. doi: 10.1016/j.neuroimage.2017.01.028. URL https://doi.org/10.1016/j.neuroimage.2017.01.028.

R. O. Duncan and G. M. Boynton. Cortical magnification within human primary visual cortex correlates with acuity thresholds. *Neuron*, 38(4):659–671, 2003. doi: 10.1016/s0896-6273(03)00265-4. URL https://doi.org/10.1016/s0896-6273(03)00265-4.

G. Edwards, P. Vetter, F. McGruer, L. S. Petro, and L. Muckli. Predictive feedback to v1 dynamically updates with sensory input. *bioRxiv*, 2017. doi: 10.1101/180539. URL https://www.biorxiv.org/content/early/2017/08/25/180539.

S. A. Engel, D. E. Rumelhart, B. A. Wandell, A. T. Lee, G. H. Glover, E.-J. Chichilnisky, and M. N. Shadlen. Fmri of human visual cortex. *Nature*, 369(6481):525–525, 1994. doi: 10.1038/369525a0. URL https://doi.org/10.1038/369525a0.

R. Epstein and N. Kanwisher. A cortical representation of the local visual environment. *Nature*, 392(6676):598–601, 1998. doi: 10.1038/33402. URL https://doi.org/10.1038/33402.

A. Falchier, S. Clavagnier, P. Barone, and H. Kennedy. Anatomical evidence of multimodal integration in primate striate cortex. *Journal of Neuroscience*, 22(13):5749–5759, 2002. ISSN 0270-6474. URL http://www.jneurosci.org/content/22/13/5749.

D. J. Felleman and D. C. Van Essen. Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, 1(1):1–47, 1991. doi: 10.1093/cercor/1.1.1. URL https://doi.org/10.1093/cercor/1.1.1.

A. Fracasso, N. Petridou, and S. O. Dumoulin. Systematic variation of population receptive field properties across cortical depth in human visual cortex. *NeuroImage*, 139(nil):427–438, 2016. doi: 10.1016/j.neuroimage.2016.06.048. URL https://doi.org/10.1016/j.neuroimage.2016.06.048.

J. Freeman, C. M. Ziemba, D. J. Heeger, E. P. Simoncelli, and J. A. Movshon. A functional and perceptual signature of the second visual area in primates. *Nat Neurosci*, 16(7):974–81, 2013. ISSN 1546-1726 (Electronic) 1097-6256 (Linking). doi: 10.1038/nn.3402. URL http://www.ncbi.nlm.nih.gov/pubmed/23685719.

K. J. Friston. Hierarchical models in the brain. *Plos Computational Biology*, 4(11), 2008. ISSN 1553-7358. doi: ARTNe1000211DOI10.1371/journal.pcbi.1000211. URL <GotoISI>://WOS:000261480800004.

P. Geladi and B. R. Kowalski. Partial least-squares regression: a tutorial. *Analytica Chimica Acta*, 185(nil):1–17, 1986. doi: 10.1016/0003-2670(86)80028-9. URL https://doi.org/10.1016/0003-2670(86)80028-9.

C. D. Gilbert and W. Li. Top-down influences on visual processing. *Nature Reviews Neuroscience*, 14(5):350–363, 2013. ISSN 1471-003X. doi: Doi10.1038/Nrn3476. URL <GotoISI>://WOS:000317913900011.

B. L. Giordano, R. A. A. Ince, J. Gross, P. G. Schyns, S. Panzeri, and C. Kayser. Contributions of local speech encoding and functional connectivity to audio-visual speech perception. *eLife*, 6(nil):nil, 2017. doi: 10.7554/elife.24763. URL https://doi.org/10.7554/elife.24763.

M. A. Goodale. Transforming vision into action. *Vision Research*, 51(13):1567–1587, 2011. doi: 10.1016/j.visres.2010.07.027. URL https://doi.org/10.1016/j.visres.2010.07.027.

M. A. Goodale and A. Milner. Separate visual pathways for perception and action. *Trends in Neurosciences*, 15(1):20–25, 1992. doi: 10.1016/0166-2236(92)90344-8. URL https://doi.org/10.1016/0166-2236(92)90344-8.

D. J. Graham, D. M. Chandler, and D. J. Field. Can the theory of "whitening" explain the center-surround properties of retinal ganglion cell receptive fields? *Vision Research*, 46 (18):2901–2913, 2006. doi: 10.1016/j.visres.2006.03.008. URL https://doi.org/10.1016/j.visres.2006.03.008.

K. Grill-Spector and R. Malach. The human visual cortex. *Annual Review of Neuroscience*, 27(1):649–677, 2004. doi: 10.1146/annurev.neuro.27.070203.144220. URL https://doi.org/10.1146/annurev.neuro.27.070203.144220.

K. Grill-Spector, T. Kushnir, S. Edelman, G. Avidan, Y. Itzchak, and R. Malach. Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron*, 24(1):187–203, 1999. doi: 10.1016/s0896-6273(00)80832-6. URL https://doi.org/10.1016/s0896-6273(00)80832-6.

K. Grill-Spector, Z. Kourtzi, and N. Kanwisher. The lateral occipital complex and its role in object recognition. *Vision Research*, 41(10-11):1409–1422, 2001. doi: 10.1016/s0042-6989(01)00073-6. URL https://doi.org/10.1016/s0042-6989(01)00073-6.

I. Groen, S. Ghebreab, V. A. Lamme, and H. S. Scholte. Spatially pooled contrast responses predict neural and perceptual similarity of naturalistic image categories. *PLoS Comput Biol*, 8(10):e1002726, 2012. ISSN 1553-7358 (Electronic) 1553-734X (Linking). doi: 10.1371/journal.pcbi.1002726. URL http://www.ncbi.nlm.nih.gov/pubmed/23093921.

I. Groen, S. Ghebreab, H. Prins, V. A. Lamme, and H. S. Scholte. From image statistics to scene gist: evoked neural activity reveals transition from low-level natural image structure to scene category. *J Neurosci*, 33(48):18814–24, 2013. ISSN 1529-2401 (Electronic) 0270-6474 (Linking). doi: 10.1523/JNEUROSCI.3128-13.2013. URL http://www.ncbi.nlm.nih.gov/pubmed/24285888.

J. Guergiuev, T. P. Lillicrap, and B. A. Richards. Towards deep learning with segregated dendrites. 2016.

O. F. Gulban and M. Schneider. Segmentator: v1.2.0, June 2017. URL https://doi.org/10.5281/zenodo.815560.

U. Güçlü and M. A. J. van Gerven. Unsupervised feature learning improves prediction of human brain activity in response to natural images. *PLoS Comput Biol*, 10(8):e1003724, 2014. doi: 10.1371/journal.pcbi.1003724.

M. Häusser and B. Mel. Dendrites: Bug or feature? *Current Opinion in Neurobiology*, 13 (3):372–383, 2003. doi: 10.1016/s0959-4388(03)00075-8. URL https://doi.org/10.1016/s0959-4388(03)00075-8.

J. V. Haxby, M. I. Gobbini, M. L. Furey, A. Ishai, J. L. Schouten, and P. Pietrini. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293(5539):2425–2430, 2001. doi: 10.1126/science.1063736. URL https://doi.org/10.1126/science.1063736.

J.-D. Haynes. A primer on pattern-based approaches to fmri: Principles, pitfalls, and perspectives. *Neuron*, 87(2):257–270, 2015. doi: 10.1016/j.neuron.2015.05.025. URL https://doi.org/10.1016/j.neuron.2015.05.025.

F. Helmchen and W. Denk. Deep tissue two-photon microscopy. *Nature Methods*, 2 (12):932–940, 2005. doi: 10.1038/nmeth818. URL https://doi.org/10.1038/nmeth818.

L. Henriksson, S. M. Khaligh-Razavi, K. Kay, and N. Kriegeskorte. Visual representations are dominated by intrinsic fluctuations correlated between areas. *Neuroimage*, 114:275–86, 2015. ISSN 1095-9572 (Electronic) 1053-8119 (Linking). doi: 10.1016/j.neuroimage.2015.04.026. URL https://www.ncbi.nlm.nih.gov/pubmed/25896934.

S. Herculano-Houzel. The human brain in numbers: a linearly scaled-up primate brain. *Frontiers in Human Neuroscience*, 3(nil):nil, 2009. doi: 10.3389/neuro.09.031.2009. URL https://doi.org/10.3389/neuro.09.031.2009.

G. Hesselmann, C. A. Kell, and A. Kleinschmidt. Ongoing activity fluctuations in hmt+ bias the perception of coherent visual motion. *Journal of Neuroscience*, 28(53):14481–14485, 2008. doi: 10.1523/jneurosci.4398-08.2008. URL https://doi.org/10.1523/jneurosci.4398-08.2008.

P. D. Hill. Kernel estimation of a distribution function. *Communications in Statistics - Theory and Methods*, 14(3):605–620, 1985. doi: 10.1080/03610928508828937. URL https://doi.org/10.1080/03610928508828937.

N. C. Hindy, F. Y. Ng, and N. B. Turk-Browne. Linking pattern completion in the hippocampus to predictive coding in visual cortex. *Nature Neuroscience*, 19(5):665–667, 2016. doi: 10.1038/nn.4284. URL https://doi.org/10.1038/nn.4284.

G. Hinton, L. Deng, D. Yu, G. Dahl, A. rahman Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. Sainath, and B. Kingsbury. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*, 29(6):82–97, 2012. doi: 10.1109/msp.2012.2205597. URL https://doi.org/10.1109/msp.2012.2205597.

D. H. Hubel. *Eye, Brain and Vision*. Scientific American Library, 1988. ISBN 0716750201. URL http://hubel.med.harvard.edu/book/bcontex.htm.

D. H. Hubel and T. N. Wiesel. Receptive fields of single neurones in the cat's striate cortex. *The Journal of Physiology*, 148(3):574–591, 1959. doi: 10.1113/jphysiol.1959.sp006308. URL https://doi.org/10.1113/jphysiol.1959.sp006308.

R. A. Ince, B. L. Giordano, C. Kayser, G. A. Rousselet, J. Gross, and P. G. Schyns. A statistical framework for neuroimaging data analysis based on mutual information estimated via a gaussian copula. *Human Brain Mapping*, 38(3):1541–1573, 2016. doi: 10.1002/hbm.23471. URL https://doi.org/10.1002/hbm.23471.

A. Inoue and L. Kilian. *Bagging Time Series Models*. Discussion Paper Series, no. 4333. Centre for Economic Policy Research, 2004. URL https://books.google.co.uk/books?id=mWQFSwAACAAJ.

E. R. Kandel. An introduction to the work of david hubel and torsten wiesel. *The Journal of Physiology*, 587(12):2733–2741, 2009. doi: 10.1113/jphysiol.2009.170688. URL https://doi.org/10.1113/jphysiol.2009.170688.

N. Kanwisher. Neuroscience: What's in a face? *Science*, 311(5761):617–618, 2006. doi: 10.1126/science.1123983. URL https://doi.org/10.1126/science.1123983.

S. Kashyap, D. Ivanov, M. Havlicek, B. A. Poser, and K. Uludağ. Impact of acquisition and analysis strategies on cortical depth-dependent fmri. *NeuroImage*, nil(nil):nil, 2017. doi: 10.1016/j.neuroimage.2017.05.022. URL https://doi.org/10.1016/j.neuroimage.2017.05.022.

K. Kay, K. Jamison, L. Vizioli, R. Zhang, E. Margalit, and K. Ugurbil. A critical assessment of data quality and venous effects in ultra-high-resolution fmri. *bioRxiv*, 2018. doi: 10.1101/337667. URL https://www.biorxiv.org/content/early/2018/06/03/337667.

K. N. Kay, T. Naselaris, R. J. Prenger, and J. L. Gallant. Identifying natural images from human brain activity. *Nature*, 452(7185):352–5, 2008. ISSN 1476-4687 (Electronic) 0028-0836 (Linking). doi: 10.1038/nature06713. URL http://www.ncbi.nlm.nih.gov/pubmed/18322462.

K. N. Kay, K. S. Weiner, and K. Grill-Spector. Attention reduces spatial uncertainty in human ventral temporal cortex. *Current Biology*, 25(5):595–600, 2015. doi: 10.1016/j.cub.2014.12.050. URL https://doi.org/10.1016/j.cub.2014.12.050.

C. Kayser. A comparison of hemodynamic and neural responses in cat visual cortex using complex stimuli. *Cerebral Cortex*, 14(8):881–891, 2004. doi: 10.1093/cercor/bhh047. URL https://doi.org/10.1093/cercor/bhh047.

H. Kennedy and J. Bullier. A double-labeling investigation of the afferent connectivity to cortical areas v1 and v2 of the macaque monkey. *J Neurosci*, 5(10):2815–30, 1985. ISSN 0270-6474 (Print) 0270-6474 (Linking). URL http://www.ncbi.nlm.nih.gov/pubmed/3840201.

K. Kessler and L. Muckli. Reading others' minds by measuring their brains: Fascinating and challenging for science, but ready for use in court? *Cortex*, 47(10):1240–1242, 2011. doi: 10.1016/j.cortex.2011.04.019. URL https://doi.org/10.1016/j.cortex.2011.04.019.

S.-M. Khaligh-Razavi and N. Kriegeskorte. Deep supervised, but not unsupervised, models may explain it cortical representation. *PLoS Computational Biology*, 10(11):e1003915, 2014. doi: 10.1371/journal.pcbi.1003915. URL https://doi.org/10.1371/journal.pcbi.1003915.

W. Klijn, B. Cumming, S. Yates, V. Karakasis, and A. Peyser. Arbor: A morphologically detailed neural network simulator for modern high performance computer architectures. 26th Computational Neuroscience Meeting, Antwerp (Belgium), 15 Jul 2017 - 20 Jul 2017, Jul 2017. URL http://juser.fz-juelich.de/record/836542.

P. Kok, L. J. Bains, T. van Mourik, D. G. Norris, and F. P. de Lange. Selective activation of the deep layers of the human primary visual cortex by top-down feedback. *Current Biology*, 26(3):371–376, 2016. doi: 10.1016/j.cub.2015.12.038. URL https://doi.org/10.1016/j.cub.2015.12.038.

D. J. Kravitz, C. S. Peng, and C. I. Baker. Real-world scene representations in high-level visual cortex: It's the spaces more than the places. *Journal of Neuroscience*, 31(20): 7322–7333, 2011. ISSN 0270-6474. doi: Doi10.1523/Jneurosci.4588-10.2011. URL <GotoISI>://WOS:000290716600011.

D. J. Kravitz, K. S. Saleem, C. I. Baker, L. G. Ungerleider, and M. Mishkin. The ventral visual pathway: an expanded neural framework for the processing of object quality. *Trends in Cognitive Sciences*, 17(1):26–49, 2013. doi: 10.1016/j.tics.2012.10.011. URL https://doi.org/10.1016/j.tics.2012.10.011.

N. Kriegeskorte, M. Mur, and P. Bandettini. Representational similarity analysis - connecting the branches of systems neuroscience. *Front Syst Neurosci*, 2:4, 2008. ISSN 1662-5137 (Electronic) 1662-5137 (Linking). doi: 10.3389/neuro.06.004.2008. URL http://www.ncbi.nlm.nih.gov/pubmed/19104670.

A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017. doi: 10.1145/3065386. URL https://doi.org/10.1145/3065386.

M. Larkum. A cellular mechanism for cortical associations: an organizing principle for the cerebral cortex. *Trends Neurosci*, 36(3):141–51, 2013. ISSN 1878-108X (Electronic) 0166-2236 (Linking). doi: 10.1016/j.tins.2012.11.006. URL https://www.ncbi.nlm.nih.gov/pubmed/23273272.

M. E. Larkum, J. J. Zhu, and B. Sakmann. A new cellular mechanism for coupling inputs arriving at different cortical layers. *Nature*, 398(6725):338–341, 1999. doi: 10.1038/18686. URL https://doi.org/10.1038/18686.

M. E. Larkum, T. Nevian, M. Sandler, A. Polsky, and J. Schiller. Synaptic integration in tuft dendrites of layer 5 pyramidal neurons: A new unifying principle. *Science*, 325(5941): 756–760, 2009. doi: 10.1126/science.1171958. URL https://doi.org/10.1126/science.1171958.

Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(7553):436–44, 2015. ISSN 1476-4687 (Electronic) 0028-0836 (Linking). doi: 10.1038/nature14539. URL http://www.ncbi.nlm.nih.gov/pubmed/26017442.

T. S. Lee. Image representation using 2d gabor wavelets. *IEEE Trans. Pattern Anal. Mach. Intell.*, 18(10):959–971, Oct. 1996. ISSN 0162-8828. doi: 10.1109/34.541406. URL http://dx.doi.org/10.1109/34.541406.

T. S. Lee and D. Mumford. Hierarchical bayesian inference in the visual cortex. *Journal of the Optical Society of America a-Optics Image Science and Vision*, 20(7):1434–1448, 2003. ISSN 0740-3232. doi: Doi10.1364/Josaa.20.001434. URL <GotoISI>://WOS:000183784000027.

T. S. Lee and M. Nguyen. Dynamics of subjective contour formation in the early visual cortex. *Proceedings of the National Academy of Sciences*, 98(4):1907–1911, 2001. doi: 10.1073/pnas.98.4.1907. URL https://doi.org/10.1073/pnas.98.4.1907.

T. S. Lee, D. Mumford, R. Romero, and V. A. F. Lamme. The role of the primary visual cortex in higher level vision. *Vision Research*, 38(15-16):2429–2454, 1998. ISSN 0042-6989. doi: Doi10.1016/S0042-6989(97)00464-1. URL <GotoISI>://WOS:000074769800013.

R. Lent, F. A. C. Azevedo, C. H. Andrade-Moraes, and A. V. O. Pinto. How many neurons do you have? some dogmas of quantitative neuroscience under revision. *European Journal of Neuroscience*, 35(1):1–9, 2011. doi: 10.1111/j.1460-9568.2011.07923.x. URL https://doi.org/10.1111/j.1460-9568.2011.07923.x.

M. D. Lescroart, D. E. Stansbury, and J. L. Gallant. Fourier power, subjective distance, and object categories all provide plausible models of bold responses in scene-selective visual areas. *Front Comput Neurosci*, 9:135, 2015. ISSN 1662-5188 (Linking). doi: 10.3389/fncom.2015.00135. URL https://www.ncbi.nlm.nih.gov/pubmed/26594164.

J. B. Levitt and J. S. Lund. The spatial extent over which neurons in macaque striate cortex pool visual signals. *Visual Neuroscience*, 19(04):439–452, 2002. doi: 10.1017/s0952523802194065. URL https://doi.org/10.1017/s0952523802194065.

M. Liang, A. Mouraux, L. Hu, and G. Iannetti. Primary sensory cortices contain distinguishable spatial patterns of activity for each sense. *Nature Communications*, 4(nil):nil, 2013. doi: 10.1038/ncomms2979. URL https://doi.org/10.1038/ncomms2979.

S. Lifshits, O. Tomer, I. Shamir, D. Barazany, G. Tsarfaty, S. Rosset, and Y. Assaf. Resolution considerations in imaging of the cortical layers. *NeuroImage*, nil(nil):nil, 2017. doi: 10.1016/j.neuroimage.2017.02.086. URL https://doi.org/10.1016/j.neuroimage.2017.02.086.

T. P. Lillicrap, D. Cownden, D. B. Tweed, and C. J. Akerman. Random synaptic feedback weights support error backpropagation for deep learning. *Nature Communications*, 7 (nil):13276, 2016. doi: 10.1038/ncomms13276. URL https://doi.org/10.1038/ncomms13276.

N. K. Logothetis. The ins and outs of fmri signals. *Nat Neurosci*, 10(10):1230–2, 2007. ISSN 1097-6256 (Print) 1097-6256 (Linking). doi: 10.1038/nn1007-1230. URL https://www.ncbi.nlm.nih.gov/pubmed/17893716.

N. K. Logothetis. What we can do and what we cannot do with fmri. *Nature*, 453(7197):869–78, 2008. ISSN 1476-4687 (Electronic) 0028-0836 (Linking). doi: 10.1038/nature06976. URL https://www.ncbi.nlm.nih.gov/pubmed/18548064.

A. H. Marblestone, G. Wayne, and K. P. Kording. Toward an integration of deep learning and neuroscience. *Frontiers in Computational Neuroscience*, 10(nil):nil, 2016. doi: 10.3389/fncom.2016.00094. URL https://doi.org/10.3389/fncom.2016.00094.

N. T. Markov and H. Kennedy. The importance of being hierarchical. *Current Opinion in Neurobiology*, 23(2):187–194, 2013. doi: 10.1016/j.conb.2012.12.008. URL https://doi.org/10.1016/j.conb.2012.12.008.

N. T. Markov, M. Ercsey-Ravasz, D. C. Van Essen, K. Knoblauch, Z. Toroczkai, and H. Kennedy. Cortical high-density counterstream architectures. *Science*, 342(6158):578–+, 2013a. ISSN 0036-8075. doi: DOI10.1126/science.1238406. URL <GotoISI>://WOS:000326334300035.

N. T. Markov, J. Vezoli, P. Chameau, A. Falchier, R. Quilodran, C. Huissoud, C. Lamy, P. Misery, P. Giroud, S. Ullman, P. Barone, C. Dehay, K. Knoblauch, and H. Kennedy. Anatomy of hierarchy: Feedforward and feedback pathways in macaque visual cortex. *Journal of Comparative Neurology*, 522(1):225–259, 2013b. doi: 10.1002/cne.23458. URL https://doi.org/10.1002/cne.23458.

N. T. Markov, M. M. Ercsey-Ravasz, A. R. Ribeiro Gomes, C. Lamy, L. Magrou, J. Vezoli, P. Misery, A. Falchier, R. Quilodran, M. A. Gariel, J. Sallet, R. Gamanut, C. Huissoud, S. Clavagnier, P. Giroud, D. Sappey-Marinier, P. Barone, C. Dehay, Z. Toroczkai, K. Knoblauch, D. C. Van Essen, and H. Kennedy. A weighted and directed interareal connectivity matrix for macaque cerebral cortex. *Cereb Cortex*, 24(1):17–36, 2014. ISSN 1460-2199 (Electronic) 1047-3211 (Linking). doi: 10.1093/cercor/bhs270. URL http://www.ncbi.nlm.nih.gov/pubmed/23010748.

F. D. Martino, E. Yacoub, V. Kemper, M. Moerel, K. Uludag, P. D. Weerd, K. Ugurbil, R. Goebel, and E. Formisano. The impact of ultra-high field mri on cognitive and computational neuroimaging. *NeuroImage*, nil(nil):nil, 2017. doi: 10.1016/j.neuroimage.2017.03.060. URL https://doi.org/10.1016/j.neuroimage.2017.03.060.

V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik,

I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, 2015. doi: 10.1038/nature14236. URL https://doi.org/10.1038/nature14236.

L. Muckli and L. S. Petro. Network interactions: non-geniculate input to v1. *Curr Opin Neurobiol*, 23(2):195–201, 2013. ISSN 1873-6882 (Electronic) 0959-4388 (Linking). doi: 10.1016/j.conb.2013.01.020. URL http://www.ncbi.nlm.nih.gov/pubmed/23402951.

L. Muckli, A. Kohler, N. Kriegeskorte, and W. Singer. Primary visual cortex activity along the apparent-motion trace reflects illusory perception. *PLoS Biol*, 3(8):e265, 2005. ISSN 1545-7885 (Electronic) 1544-9173 (Linking). doi: 10.1371/journal.pbio.0030265. URL http://www.ncbi.nlm.nih.gov/pubmed/16018720.

L. Muckli, L. S. Petro, and F. W. Smith. Backwards is the way forward: feedback in the cortical hierarchy predicts the expected future. *Behav Brain Sci*, 36(3):221, 2013. ISSN 1469-1825 (Electronic) 0140-525X (Linking). doi: 10.1017/S0140525X12002361. URL http://www.ncbi.nlm.nih.gov/pubmed/23663531.

L. Muckli, F. De Martino, L. Vizioli, L. S. Petro, F. W. Smith, K. Ugurbil, R. Goebel, and E. Yacoub. Contextual feedback to superficial layers of v1. *Curr Biol*, 25(20):2690–5, 2015. ISSN 1879-0445 (Electronic) 0960-9822 (Linking). doi: 10.1016/j.cub.2015.08.057. URL http://www.ncbi.nlm.nih.gov/pubmed/26441356.

D. Mumford. On the computational architecture of the neocortex .1. the role of the thalamocortical loop. *Biological Cybernetics*, 65(2):135–145, 1991. ISSN 0340-1200. doi: Doi10.1007/Bf00202389. URL <GotoISI>://WOS:A1991FW17300008.

D. Mumford. On the computational architecture of the neocortex .2. the role of cortico-cortical loops. *Biological Cybernetics*, 66(3):241–251, 1992. ISSN 0340-1200. doi: Doi10.1007/Bf00198477. URL <GotoISI>://WOS:A1992GZ45400004.

D. Mumford and A. Desolneux. *Pattern theory: the stochastic analysis of real-world signals*. CRC Press, 2010.

T. Naselaris, K. N. Kay, S. Nishimoto, and J. L. Gallant. Encoding and decoding in fmri. *NeuroImage*, 56(2):400–410, 2011. doi: 10.1016/j.neuroimage.2010.07.073. URL https://doi.org/10.1016/j.neuroimage.2010.07.073.

T. Naselaris, C. A. Olman, D. E. Stansbury, K. Ugurbil, and J. L. Gallant. A voxel-wise encoding model for early visual areas decodes mental images of remembered scenes. *Neuroimage*, 105:215–28, 2015. ISSN 1095-9572 (Electronic) 1053-8119 (Linking). doi: 10.1016/j.neuroimage.2014.10.018. URL http://www.ncbi.nlm.nih.gov/pubmed/25451480.

R. Naud and H. Sprekeler. Burst ensemble multiplexing: A neural code connecting dendritic spikes with microcircuits. *bioRxiv*, 2017. doi: 10.1101/143636. URL https://www.biorxiv.org/content/early/2017/06/30/143636.

H. Nili, C. Wingfield, A. Walther, L. Su, W. Marslen-Wilson, and N. Kriegeskorte. A toolbox for representational similarity analysis. *PLoS Comput Biol*, 10(4):e1003553, 2014. ISSN 1553-7358 (Electronic) 1553-734X (Linking). doi: 10.1371/journal.pcbi.1003553. URL http://www.ncbi.nlm.nih.gov/pubmed/24743308.

A. Oliva and A. Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *Int. J. Comput. Vision*, 42(3):145–175, 2001. ISSN 0920-5691. doi: 10.1023/a:1011139631724. URL http://link.springer.com/article/10.1023%2FA%3A1011139631724.

A. Oliva and A. Torralba. Building the gist of a scene: the role of global image features in recognition. *Visual Perception, Pt 2: Fundamentals of Awareness: Multi-Sensory Integration and High-Order Perception*, 155:23–36, 2006. ISSN 0079-6123. doi: 10.1016/S0079-6123(06)55002-2. URL <GotoISI>://WOS:000245615200002.

C. A. Olman, N. Harel, D. A. Feinberg, S. He, P. Zhang, K. Ugurbil, and E. Yacoub. Layer-specific fmri reflects different neuronal computations at different depths in human v1. *PLoS ONE*, 7(3):e32536, 2012. doi: 10.1371/journal.pone.0032536. URL https://doi.org/10.1371/journal.pone.0032536.

B. A. Olshausen and D. J. Field. Sparse coding with an overcomplete basis set: A strategy employed by v1? *Vision Research*, 37(23):3311–3325, 1997. ISSN 0042-6989. doi: Doi10.1016/S0042-6989(97)00169-7. URL <GotoISI>://WOS:A1997YG15500010.

B. A. Olshausen and D. J. Field. How close are we to understanding v1? *Neural Computation*, 17(8):1665–1699, 2005. doi: 10.1162/0899766054026639. URL https://doi.org/10.1162/0899766054026639.

L. S. Petro and L. Muckli. The laminar integration of sensory inputs with feedback signals in human cortex. *Brain and Cognition*, 112(nil):54–57, 2017. doi: 10.1016/j.bandc.2016.06.007. URL https://doi.org/10.1016/j.bandc.2016.06.007.

L. S. Petro, F. W. Smith, P. G. Schyns, and L. Muckli. Decoding face categories in diagnostic subregions of primary visual cortex. *Eur J Neurosci*, 37(7):1130–9, 2013. ISSN 1460-9568 (Electronic) 0953-816X (Linking). doi: 10.1111/ejn.12129. URL http://www.ncbi.nlm.nih.gov/pubmed/23373719.

L. S. Petro, L. Vizioli, and L. Muckli. Contributions of cortical feedback to sensory processing in primary visual cortex. *Front Psychol*, 5:1223, 2014. ISSN 1664-1078 (Electronic)

1664-1078 (Linking). doi: 10.3389/fpsyg.2014.01223. URL http://www.ncbi.nlm.nih.gov/pubmed/25414677.

W. A. Phillips, A. Clark, and S. M. Silverstein. On the functions, mechanisms, and malfunctions of intracortical contextual modulation. *Neurosci Biobehav Rev*, 52:1–20, 2015. ISSN 1873-7528 (Electronic) 0149-7634 (Linking). doi: 10.1016/j.neubiorev.2015.02.010. URL http://www.ncbi.nlm.nih.gov/pubmed/25721105.

M. E. Raichle. The restless brain. *Brain Connectivity*, 1(1):3–12, 2011. doi: 10.1089/brain.2011.0019. URL https://doi.org/10.1089/brain.2011.0019.

M. E. Raichle and M. A. Mintun. Brain work and brain imaging. *Annual Review of Neuroscience*, 29(1):449–476, 2006. doi: 10.1146/annurev.neuro.29.051605.112819. URL https://doi.org/10.1146/annurev.neuro.29.051605.112819.

R. P. N. Rao and D. H. Ballard. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1):79–87, 1999. doi: 10.1038/4580. URL https://doi.org/10.1038/4580.

D. Reichert and T. Serre. Neuronal synchrony in complex-valued deep networks. *arXiv preprint arXiv:1312.6115*, 2013.

K. S. Rockland and G. W. V. Hoesen. Direct temporal-occipital feedback connections to striate cortex (v1) in the macaque monkey. *Cerebral Cortex*, 4(3):300–313, 1994. doi: 10.1093/cercor/4.3.300. URL https://doi.org/10.1093/cercor/4.3.300.

K. S. Rockland and H. Ojima. Multisensory convergence in calcarine visual areas in macaque monkey. *International Journal of Psychophysiology*, 50(1-2):19–26, 2003. doi: 10.1016/s0167-8760(03)00121-1. URL https://doi.org/10.1016/s0167-8760(03)00121-1.

K. S. Rockland, K. S. Saleem, and K. Tanaka. Divergent feedback connections from areas v4 and teo in the macaque. *Vis Neurosci*, 11(3):579–600, 1994. ISSN 0952-5238 (Print) 0952-5238 (Linking). URL https://www.ncbi.nlm.nih.gov/pubmed/8038130.

H. S. Scholte, S. Ghebreab, L. Waldorp, A. W. M. Smeulders, and V. a. F. Lamme. Brain responses strongly correlate with weibull image statistics when processing natural images. *Journal of Vision*, 9:1–15, 2009. doi: 10.1167/9.4.29.Introduction. URL http://www.journalofvision.org/content/9/4/29.full.pdf.

P. G. Schyns and A. Oliva. From blobs to boundary edges - evidence for time-scale-dependent and spatial-scale-dependent scene recognition. *Psychological Science*, 5(4):195–200, 1994. ISSN 0956-7976. doi: DOI10.1111/j.1467-9280.1994.tb00500.x. URL <GotoISI>://WOS:A1994PB01600007.

D. W. Scott. *Multivariate Density Estimation: Theory, Practice, and Visualization*. Wiley Series in Probability and Statistics. Wiley, 2009. ISBN 9780470317686. URL https://books.google.co.uk/books?id=hE5A3ZYXc14C.

M. Sereno, A. Dale, J. Reppas, K. Kwong, J. Belliveau, T. Brady, B. Rosen, and R. Tootell. Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science*, 268(5212):889–893, 1995. doi: 10.1126/science.7754376. URL https://doi.org/10.1126/science.7754376.

M. I. Sereno. Brain mapping in animals and humans. *Current Opinion in Neurobiology*, 8 (2):188–194, 1998. doi: 10.1016/s0959-4388(98)80139-6. URL https://doi.org/10.1016/s0959-4388(98)80139-6.

T. Serre, A. Oliva, and T. Poggio. A feedforward architecture accounts for rapid categorization. *Proc Natl Acad Sci U S A*, 104(15):6424–9, 2007. ISSN 0027-8424 (Print) 0027-8424 (Linking). doi: 10.1073/pnas.0700622104. URL http://www.ncbi.nlm.nih.gov/pubmed/17404214.

Z. Shao and A. Burkhalter. Different balance of excitation and inhibition in forward and feedback circuits of rat visual cortex. *Journal of Neuroscience*, 16(22):7353–7365, 1996. ISSN 0270-6474. URL http://www.jneurosci.org/content/16/22/7353.

R. K. Shepherd, M. N. Shivdasani, D. A. Nayagam, C. E. Williams, and P. J. Blamey. Visual prostheses for the blind. *Trends in Biotechnology*, 31(10):562–571, 2013. doi: 10.1016/j.tibtech.2013.07.001. URL https://doi.org/10.1016/j.tibtech.2013.07.001.

B. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis, 1986. ISBN 9780412246203. URL https://books.google.co.uk/books?id=e-xsrjsL7WkC.

L. C. Sincich and J. C. Horton. The circuitry of v1 and v2: Integration of color, form, and motion. *Annual Review of Neuroscience*, 28:303–326, 2005. ISSN 0147-006X. doi: DOI10.1146/annurev.neuro.28.061604.135731. URL <GotoISI>://WOS:000231235700012.

F. Smith, L. Petro, P. Schyns, and L. Muckli. Complex contextual processing in v1 during face categorizations. *Journal of Vision*, 10(7):657–657, 2010. doi: 10.1167/10.7.657. URL https://doi.org/10.1167/10.7.657.

F. W. Smith and L. Muckli. Nonstimulated early visual areas carry information about surrounding context. *Proc Natl Acad Sci USA*, 107(46):20099–103, 2010. ISSN 1091-6490 (Electronic) 0027-8424 (Linking). doi: 10.1073/pnas.1000233107. URL http://www.ncbi.nlm.nih.gov/pubmed/21041652.

S. M. Smith, M. Jenkinson, M. W. Woolrich, C. F. Beckmann, T. E. Behrens, H. Johansen-Berg, P. R. Bannister, M. D. Luca, I. Drobnjak, D. E. Flitney, R. K. Niazy, J. Saunders, J. Vickers, Y. Zhang, N. D. Stefano, J. M. Brady, and P. M. Matthews. Advances in functional and structural mr image analysis and implementation as fsl. *NeuroImage*, 23 (nil):S208–S219, 2004. doi: 10.1016/j.neuroimage.2004.07.051. URL https://doi.org/10.1016/j.neuroimage.2004.07.051.

C. Spoerer, P. McClure, and N. Kriegeskorte. Recurrent convolutional neural networks: A better model of biological object recognition under occlusion. *bioRxiv*, 2017.

M. W. Spratling. Reconciling predictive coding and biased competition models of cortical function. *Front Comput Neurosci*, 2:4, 2008. ISSN 1662-5188 (Electronic) 1662-5188 (Linking). doi: 10.3389/neuro.10.004.2008. URL http://www.ncbi.nlm.nih.gov/pubmed/18978957.

Y. Sugita. Grouping of image fragments in primary visual cortex. *Nature*, 401(6750):269–72, 1999. ISSN 0028-0836 (Print) 0028-0836 (Linking). doi: 10.1038/45785. URL http://www.ncbi.nlm.nih.gov/pubmed/10499583.

N. Takahashi, T. G. Oertner, P. Hegemann, and M. E. Larkum. Active cortical dendrites modulate perception. *Science*, 354(6319):1587–1590, 2016. ISSN 1095-9203 (Electronic) 0036-8075 (Linking). doi: 10.1126/science.aah6066. URL https://www.ncbi.nlm.nih.gov/pubmed/28008068.

C. O. Tan. Anticipatory changes in regional cerebral hemodynamics: A new role for dopamine? *Journal of Neurophysiology*, 101(6):2738–2740, 2009. doi: 10.1152/jn.00141.2009. URL https://doi.org/10.1152/jn.00141.2009.

K. Tanaka. Inferotemporal cortex and object vision. *Annual Review of Neuroscience*, 19 (1):109–139, 1996. doi: 10.1146/annurev.ne.19.030196.000545. URL https://doi.org/10.1146/annurev.ne.19.030196.000545.

P. Tian, I. C. Teng, L. D. May, R. Kurz, K. Lu, M. Scadeng, E. M. C. Hillman, A. J. D. Crespigny, H. E. D'Arceuil, J. B. Mandeville, J. J. A. Marota, B. R. Rosen, T. T. Liu, D. A. Boas, R. B. Buxton, A. M. Dale, and A. Devor. Cortical depth-specific microvascular dilation underlies laminar differences in blood oxygenation level-dependent functional mri signal. *Proceedings of the National Academy of Sciences*, 107(34):15246–15251, 2010. doi: 10.1073/pnas.1006735107. URL https://doi.org/10.1073/pnas.1006735107.

R. B. H. Tootell, N. K. Hadjikhani, W. Vanduffel, A. K. Liu, J. D. Mendola, M. I. Sereno, and A. M. Dale. Functional analysis of primary visual cortex (v1) in humans. *Proceedings of the National Academy of Sciences*, 95(3):811–817, 1998. doi: 10.1073/pnas.95.3.811. URL https://doi.org/10.1073/pnas.95.3.811.

A. Torralba and A. Oliva. Depth estimation from image structure. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(9):1226–1238, 2002. ISSN 0162-8828. doi: 10.1109/tpami.2002.1033214.

R. Trampel, P.-L. Bazin, K. Pine, and N. Weiskopf. In-vivo magnetic resonance imaging (mri) of laminae in the human cortex. *NeuroImage*, nil(nil):nil, 2017. doi: 10.1016/j.neuroimage.2017.09.037. URL https://doi.org/10.1016/j.neuroimage.2017.09.037.

R. Turner, P. Jezzard, H. Wen, K. K. Kwong, D. L. Bihan, T. Zeffiro, and R. S. Balaban. Functional mapping of the human visual cortex at 4 and 1.5 tesla using deoxygenation contrast epi. *Magnetic Resonance in Medicine*, 29(2):277–279, 1993. doi: 10.1002/mrm.1910290221. URL https://doi.org/10.1002/mrm.1910290221.

K. Ugurbil. Magnetic resonance imaging at ultrahigh fields. *IEEE Transactions on Biomedical Engineering*, 61(5):1364–1379, 2014. doi: 10.1109/tbme.2014.2313619. URL https://doi.org/10.1109/tbme.2014.2313619.

K. Uludağ and P. Blinder. Linking brain vascular physiology to hemodynamic response in ultra-high field mri. *NeuroImage*, nil(nil):nil, 2017. doi: 10.1016/j.neuroimage.2017.02.063. URL https://doi.org/10.1016/j.neuroimage.2017.02.063.

J. Vaughan, M. Garwood, C. Collins, W. Liu, L. DelaBarre, G. Adriany, P. Andersen, H. Merkle, R. Goebel, M. Smith, and K. Ugurbil. 7t vs. 4t: Rf power, homogeneity, and signal-to-noise comparison in head images. *Magnetic Resonance in Medicine*, 46(1):24–30, 2001. doi: 10.1002/mrm.1156. URL https://doi.org/10.1002/mrm.1156.

P. Vetter, M. H. Grosbras, and L. Muckli. Tms over v5 disrupts motion prediction. *Cereb Cortex*, 2013. ISSN 1460-2199 (Electronic) 1047-3211 (Linking). doi: 10.1093/cercor/bht297. URL http://www.ncbi.nlm.nih.gov/pubmed/24152544.

P. Vetter, F. W. Smith, and L. Muckli. Decoding sound and imagery content in early visual cortex. *Current Biology*, 24(11):1256–1262, 2014. ISSN 0960-9822. doi: DOI10.1016/j.cub.2014.04.020. URL <GotoISI>://WOS:000336668500027.

J. D. Victor, K. Purpura, E. Katz, and B. Mao. Population encoding of spatial frequency, orientation, and color in macaque v1. *Journal of Neurophysiology*, 72(5):2151–2166, 1994. doi: 10.1152/jn.1994.72.5.2151. URL https://doi.org/10.1152/jn.1994.72.5.2151.

A. Walther, H. Nili, N. Ejaz, A. Alink, N. Kriegeskorte, and J. Diedrichsen. Reliability of dissimilarity measures for multi-voxel pattern analysis. *Neuroimage*, 137:188–200, 2016. ISSN 1095-9572 (Electronic) 1053-8119 (Linking). doi: 10.1016/j.neuroimage.2015.12.012. URL https://www.ncbi.nlm.nih.gov/pubmed/26707889.

D. B. Walther, E. Caddigan, L. Fei-Fei, and D. M. Beck. Natural scene categories revealed in distributed patterns of activity in the human brain. *Journal of Neuroscience*, 29(34): 10573–10581, 2009. ISSN 0270-6474. doi: Doi10.1523/Jneurosci.0559-09.2009. URL <GotoISI>://WOS:000269317900012.

Y. Wang, S. Celebrini, Y. Trotter, and P. Barone. Visuo-auditory interactions in the primary visual cortex of the behaving monkey: Electrophysiological evidence. *BMC Neuroscience*, 9(1):79, 2008. doi: 10.1186/1471-2202-9-79. URL https://doi.org/10.1186/1471-2202-9-79.

H. Wässle. Parallel processing in the mammalian retina. *Nature Reviews Neuroscience*, 5 (10):747–757, 2004. doi: 10.1038/nrn1497. URL https://doi.org/10.1038/nrn1497.

G. Westheimer. Center-surround antagonism in spatial vision: Retinal or cortical locus? *Vision Research*, 44(21):2457–2465, 2004. doi: 10.1016/j.visres.2004.05.014. URL https://doi.org/10.1016/j.visres.2004.05.014.

M. A. Williams, C. I. Baker, H. P. Op de Beeck, W. M. Shim, S. Dang, C. Triantafyllou, and N. Kanwisher. Feedback of visual object information to foveal retinotopic cortex. *Nat Neurosci*, 11(12):1439–45, 2008. ISSN 1546-1726 (Electronic) 1097-6256 (Linking). doi: 10.1038/nn.2218. URL http://www.ncbi.nlm.nih.gov/pubmed/18978780.

J. Xiao, J. Hays, K. A. Ehinger, A. Oliva, and A. Torralba. Sun database: Large-scale scene recognition from abbey to zoo. In *CVPR*, pages 3485–3492. IEEE Computer Society, 2010. ISBN 978-1-4244-6984-0.

E. Yacoub, N. Harel, and K. Ugurbil. High-field fmri unveils orientation columns in humans. *Proceedings of the National Academy of Sciences*, 105(30):10607–10612, 2008. doi: 10.1073/pnas.0804110105. URL https://doi.org/10.1073/pnas.0804110105.

U. Yarach, M.-H. In, I. Chatnuntawech, B. Bilgic, F. Godenschweger, H. Mattern, A. Sciarra, and O. Speck. Model-based iterative reconstruction for single-shot epi at 7t. *Magnetic Resonance in Medicine*, nil(nil):nil, 2017. doi: 10.1002/mrm.26633. URL https://doi.org/10.1002/mrm.26633.

C. C.-C. Yen, D. Papoti, and A. C. Silva. Investigating the spatiotemporal characteristics of the deoxyhemoglobin-related and deoxyhemoglobin-unrelated functional hemodynamic response across cortical layers in awake marmosets. *NeuroImage*, nil(nil):nil, 2017. doi: 10.1016/j.neuroimage.2017.03.005. URL https://doi.org/10.1016/j.neuroimage.2017.03.005.

X. Yu, C. Qian, D. yow Chen, S. J. Dodd, and A. P. Koretsky. Deciphering laminar-specific neural inputs with line-scanning fmri. *Nature Methods*, 11(1):55–58, 2013. doi: 10.1038/nmeth.2730. URL https://doi.org/10.1038/nmeth.2730.

P. A. Yushkevich, Y. Gao, and G. Gerig. Itk-snap: An interactive tool for semi-automatic segmentation of multi-modality biomedical images. In *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the*, pages 3342–3345. IEEE, 2016.

G. Zarkadakis. *In Our Own Image: Will artificial intelligence save or destroy us?*. Random House, 2015. ISBN 1846044367.