

# **Shape Processing across Lateral Occipital Cortex**

Richard Jeffrey William Vernon

PhD

University of York

Psychology

November 2016

## Abstract

There are two predominant means of identifying visual areas in the human brain; retinotopy (exploiting maps of the visual field) and localisers (exploiting functional selectivity). This thesis aimed to bridge those two approaches, assessing the roles of LO-1 and LO-2; two retinotopically-defined regions that show overlap with the functionally-defined (shape selective) Lateral Occipital Complex (LOC). More generally, we asked what is the nature of the shape representation across Lateral Occipital cortex?

We first probed the functional roles of LO-1 and LO-2, finding that LO-2 is the more shape-sensitive region of the pair and will respond to second order shape stimuli, whereas LO-1 may process more local cues (perhaps orientation information).

Our later work then assessed neural shape representations across visual cortex, identifying two discrete representations; ‘Shape-profile’ (essentially retinotopic responses) and ‘Shape-complexity’ (responses based upon the complexity of a shape’s contour). The latter dimension captured variance in LOC, and surprisingly LO-2. This indicates that even explicit visual field maps can respond to non-retinotopic attributes such as curvature complexity. Intriguingly, a transition between dimensions occurred around LO-1 and LO-2.

Finally, we explicitly tested whether the ‘Shape-complexity’ representation may be curvature based. Our results implied that radial shape protrusions are highly salient features for Lateral Occipital cortex, but it is not necessarily the points of maximal curvature that are being responded to. Instead, we hypothesise that it is the convergent lines comorbid with curvature that neurons may be attuned to, as such lines likely represent the most salient or characteristic features in a given shape.

In sum, we argue for an evolving shape representation across visual cortex, with some degree of shape sensitivity first emerging around LO-1 and LO-2. These maps may then be acting as preliminary processing stages for more selective shape tunings in LOC.

# List of Contents

Abstract .....	2
List of Contents .....	3
List of Tables.....	9
List of Figures .....	10
Acknowledgements .....	14
Declaration .....	15
Chapter 1. Introduction .....	16
1.1 Overview .....	16
1.2 The Lateral Occipital Complex (LOC) .....	17
1.2.1 Object recognition and the brain .....	17
1.3 Retinotopy .....	23
1.3.1 The travelling-wave method .....	23
1.3.2 Early visual cortex.....	25
1.3.3 LO-1 and LO-2.....	28
1.3.4 Retinotopic organisation of the Lateral Occipital Cortex .....	30
1.3.5 Additional properties of LO-1 and LO-2 .....	32
1.4 Properties of LOC .....	34
1.4.1 Section overview .....	34
1.4.2 Cue Invariance .....	35
1.4.3 Boundaries and Contours .....	38
1.4.4 Perceptual Experience of Shape.....	47
1.5 Concluding Remarks .....	53
Chapter 2. Methods Review .....	55
2.1 Overview .....	55
2.2 Functional Magnetic Resonance Imaging (fMRI).....	56
2.2.1 The BOLD response.....	56

2.2.2	Measuring the BOLD response.....	58
2.2.3	Scan Parameters .....	60
2.3	Structural Scans .....	62
2.3.1	Data acquisition.....	62
2.3.2	Data Analysis .....	64
2.3.3	Brain Segmentation.....	66
2.3.4	Axial PDs .....	66
2.4	Experimental Designs.....	67
2.4.1	Block Designs .....	67
2.4.2	Rapid Event-related Designs.....	69
2.4.3	Additional Design Considerations .....	70
2.4.4	Data Analysis I.....	71
2.4.5	Data Analysis II.....	74
2.4.6	Data Analysis III .....	76
2.5	LOC Functional Localiser .....	76
2.5.1	Defining the Lateral Occipital Complex (LOC) .....	78
2.5.2	Stimulus Generation.....	80
2.5.3	Stimulus Presentation.....	81
2.5.4	Data Analysis .....	83
2.6	Retinotopic Mapping .....	85
2.6.1	Stimulus Generation.....	85
2.6.2	Stimulus Presentation.....	86
2.6.3	Stimulus Presentation – Future Considerations .....	87
2.6.4	Data Analysis .....	89
2.6.5	Identifying Regions of Interest (ROIs) .....	91
2.7	Radial Frequency (RF) patterns .....	93
2.8	Fourier Descriptors (FDs).....	96

2.9	Summary .....	99
Chapter 3. Responses to Orientation and Shape Processing Tasks in Retinotopic Regions LO-1 and LO-2.....		
		100
3.1	Abstract .....	100
3.2	Introduction .....	101
3.3	Methods .....	103
3.3.1	Participants .....	103
3.3.2	Psychophysics .....	103
3.3.3	Functional MRI .....	106
3.4	Results .....	107
3.4.1	Behavioural Analysis .....	107
3.4.2	Hemispheric differences.....	108
3.4.3	ROI analysis .....	109
3.5	Discussion .....	111
Chapter 4. Responses to Shape from Form Coherence in the Lateral Occipital Cortex .....		
		115
4.1	Abstract .....	115
4.2	Introduction .....	116
4.3	Methods .....	117
4.3.1	Participants.....	117
4.3.2	Stimulus Presentation.....	118
4.3.3	Functional MRI data .....	118
4.4	Data Analysis .....	120
4.4.1	Retinotopy .....	120
4.4.2	Functional Data .....	121
4.4.3	Localiser .....	121
4.5	Results .....	122
4.5.1	Responses in early retinotopic visual areas V1-V4.....	122

4.5.2	Responses in Lateral Occipital retinotopic visual areas LO-1 and LO-2	123
4.5.3	Responses in functionally-defined visual areas LO and pFs .....	124
4.5.4	Whole-brain analysis.....	125
4.6	Discussion .....	126
Chapter 5. Multivariate Patterns in the Human Object-Processing Pathway Reveal a Shift from Retinotopic to Shape Curvature Representations in LO-1 and LO-2.....		
5.1	Abstract .....	129
5.2	Introduction .....	130
5.3	Materials and Methods .....	132
5.3.1	Participants.....	132
5.3.2	Preliminary Data Acquisition and Analysis.....	132
5.3.3	Main Functional Scans: Stimuli.....	134
5.3.4	Main Functional Scans: Data Acquisition and Analysis.....	137
5.3.5	ROI Restriction .....	138
5.3.6	Correlation Analysis and Shape Similarity Metrics.....	139
5.3.7	Perceptual Similarity Measure .....	140
5.3.8	Multidimensional Scaling (MDS).....	141
5.3.9	Statistical Analysis.....	142
5.4	Results .....	142
5.4.1	Correlation Analysis.....	143
5.4.2	Result Reliability.....	146
5.4.3	Multidimensional Scaling (MDS).....	150
5.5	Discussion .....	152
Chapter 6. Testing for Linearity in Fourier Descriptor ‘Information’ Content.....		
6.1	Abstract .....	157
6.2	Introduction .....	158
6.3	Methods.....	160

6.3.1	Participants .....	160
6.3.2	Stimuli Creation .....	160
6.3.3	Stimulus Presentation.....	163
6.3.4	Predicting linearity .....	165
6.3.5	Statistical Analysis .....	167
6.4	Results .....	167
6.4.1	Counterbalancing Measures .....	167
6.4.2	Stimuli Results .....	168
6.4.3	Predicting midpoint judgements .....	170
6.5	Discussion .....	172
Chapter 7. Exploring the Role of Curvature for Neural Representations within		
	Shape-Selective Lateral Occipital Cortex .....	176
7.1	Abstract .....	176
7.2	Introduction .....	177
7.3	General Methods .....	179
7.3.1	Participants.....	179
7.3.2	Preliminary Data Acquisition and Analysis.....	180
7.3.3	ROI Restriction .....	180
7.4	Methods - Radial Frequency Experiment (Expt. 1).....	182
7.4.1	Stimulus Creation.....	182
7.4.2	Data Acquisition and Analysis.....	184
7.4.3	Correlation Analysis and Shape Similarity Metrics.....	185
7.4.4	Visualising Predictors .....	188
7.5	Results - Radial Frequency Experiment (Expt. 1).....	190
7.5.1	Stimulus Size (Figure 7.4).....	190
7.5.2	Amplitude and Frequency (Figure 7.5).....	191
7.5.3	Additional influences (Figure 7.6).....	191

7.5.4	Extending the ventral stream; VO-1/2 .....	193
7.6	Discussion - Radial Frequency Experiment (Expt. 1).....	193
7.7	Methods – Gabor Element Experiment (Expt. 2).....	196
7.7.1	Stimulus Creation.....	196
7.7.2	Data Acquisition and Analysis.....	199
7.8	Results – Gabor Element Experiment (Expt. 2).....	200
7.8.1	‘Not Curves’ versus ‘Curves’ – concentricity?.....	202
7.9	Discussion – Gabor Element Experiment (Expt. 2) .....	203
7.10	General Discussion.....	206
Chapter 8.	General Discussion.....	209
8.1	Overview .....	209
8.2	Summary of Work.....	209
8.3	Interpretation of the Results .....	212
8.3.1	LO-1 .....	212
8.3.2	LO-2 .....	215
8.3.3	LO .....	218
8.4	Considering a protrusion-based representation .....	223
8.5	Macaque V4 .....	226
8.6	Concluding remarks .....	228
References	.....	230

## List of Tables

**Table 5.1** Percentage overlap between LO-1, LO-2 and LO Localiser data across all participants

**Table 5.2.** Comparing shape factor/neural similarity correlations with and without spatial smoothing

## List of Figures

- Figure 1.1.** The Lateral Occipital Complex
- Figure 1.2.** The Lincoln illusion
- Figure 1.3.** Retinotopic Mapping
- Figure 1.4.** Receptive fields in early visual cortex
- Figure 1.5.** Region of interest overview
- Figure 1.6.** Population receptive field (pRF) sizes
- Figure 1.7.** LO-1/LO-2 double dissociation
- Figure 1.8.** Objects from motion and texture
- Figure 1.9.** Greyscale and line drawing objects
- Figure 1.10.** 2D versus 3D stimuli
- Figure 1.11.** Shape versus contours
- Figure 1.12.** Contour completion
- Figure 1.13.** Deleting local features versus parts
- Figure 1.14.** Kanizsa-style stimuli
- Figure 1.15.** Objects versus holes
- Figure 1.16.** Novel object stimuli
- Figure 1.17.** Neural similarity in LOC
- Figure 1.18.** Perceptual versus physical shape similarity
- Figure 2.1.** Double gamma hemodynamic response function (HRF)
- Figure 2.2.** Predicting neural responses
- Figure 2.3.** Example Anatomical Images
- Figure 2.4.** Fourier phase scrambling

- Figure 2.5.** Object Scrambling Process
- Figure 2.6.** Retinotopic Mapping Stimuli
- Figure 2.7.** Travelling Wave Analysis
- Figure 2.8.** Example Retinotopic Maps
- Figure 2.9.** Example Radial Frequency (RF) Patterns
- Figure 2.10.** Fourth Derivative of a Gaussian (D4)
- Figure 2.11.** D4 Rendering Method
- Figure 2.12.** Fourier Descriptors (FDs)
- Figure 3.1.** Example stimuli
- Figure 3.2.** Example sigmoid fits for one participant
- Figure 3.3.** Summary of Results
- Figure 4.1.** Example second-order shape stimulus
- Figure 4.2.** Regions of interest and functional activity for one representative participant
- Figure 4.3.** Percent signal change for different brain regions
- Figure 5.1.** Identification of visual areas
- Figure 5.2.** Stimuli used in the study
- Figure 5.3.** Percent signal change for different brain regions
- Figure 5.4.** Correlations between predictors, factors and neural similarity
- Figure 5.5.** Group mean neural similarity matrices and their correlation with 'Shape-Profile' and 'Shape-Complexity' factors
- Figure 5.6.** Solutions from multidimensional scaling
- Figure 6.1.** Choice of linear transitions

- Figure 6.2.** Target Fourier Descriptor (FD)
- Figure 6.3.** Predicting FD amplitude
- Figure 6.4.** Example stimuli
- Figure 6.5.** Image comparison order
- Figure 6.6.** Average participant linearity choices
- Figure 6.7.** Participant versus Linear responses
- Figure 6.8.** Multidimensional Scaling solution for participant's midpoint judgements
- Figure 6.9.** Comparing the predictive power of stimulus characteristics to explain participant midpoint judgements
- Figure 6.10.** Predicted versus actual amplitude
- Figure 7.1.** Radial Frequency (RF) Stimuli
- Figure 7.2.** Correlations between RF pattern perimeter and RF pattern area with neural similarity
- Figure 7.3.** Sample lobe
- Figure 7.4.** RF predictor variables
- Figure 7.5.** Correlations between RF pattern amplitude and RF pattern frequency with neural similarity
- Figure 7.6.** Correlations between the 'Lobe prominence' and 'Lobe curvature' factors with neural similarity
- Figure 7.7.** Example Gabor Element Stimuli
- Figure 7.8.** Percent signal change across early visual cortex (V1-V3 plus V4) for each condition
- Figure 7.9.** Percent signal change across Lateral Occipital ROIs (LO-1/2, LO) plus pFs for each condition

**Figure 8.1.** Hypothetical protrusion detection

**Figure 8.2.** Deleting convergent lines

## Acknowledgements

I would first like to offer sincere thanks and gratitude to my supervisor, Antony Morland. Throughout the past three years he has offered excellent guidance and encouragement, which undoubtedly helped to develop my fledgling career in science. He also showed considerable patience towards my occasional (!) scepticism and pedantic attention to detail, whilst ensuring that our lab maintained an enjoyable and scientifically engaging atmosphere.

I would also like to thank the members of my thesis advisory panel; Alex Wade and Heidi Baseler. Their advice and constructive criticism certainly elevated the quality of the work contained in this thesis. I also thank André Gouws, who was always ready to battle a particularly tricky alignment, help debug fiendishly fickle code, or buy a round in on a Thursday.

I would next like to thank my parents for their support throughout the years. In particular, I thank them for supporting me through the degree change that led to my pursuit of Psychology, and ultimately this PhD. I also thank my sister for her encouragement, and my Nan, who helped fund the first year of my Psychology degree, and always lauded the importance of education for personal development.

I would also like to offer thanks to my friends and colleagues in the department of Psychology, and the York Neuroimaging Centre. Particularly, I thank Samuel Lawrence, Holly Brown and Rachel Woodall. They sometimes say a PhD can be a lonely affair, but with their constant friendship it has been anything but. The times we have shared over the last three years, whilst leading to a few... 'compromising' pictures, have nevertheless been immense fun. They have certainly made this PhD one of the more enjoyable periods of my life.

Last, and certainly least, I would like to thank my cat for occasionally acknowledging my existence, Mendeley for saving me from hand-editing countless citations, my computer all those times it didn't immediately crash when I suddenly realised I hadn't saved my work for a few hours, the original Rabbit for proving so amenable to Fourier analysis, the Rook and Gaskill for many great Thursdays (and slow Fridays), and the bananas, coffee and occasional Browns sandwich that fuelled the working day.

## Declaration

I declare that this thesis is a presentation of original work, carried out under the supervision of Professor Antony B. Morland, and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as references.

Data reported in **Chapter 3**, plus the retinotopic mapping data used throughout this thesis were collected by myself and another PhD researcher; Samuel J D Lawrence. All data were analysed, interpreted and written-up independently for the purpose of this thesis.

Data reported in **Chapter 3** were presented at the Society for Neuroscience conference (2014), Washington, USA, as:

*Vernon, RJW, Lawrence, SJD, Gouws, AD, Morland, AB (2014) Responses to Shape from Form Coherence in the Lateral Occipital Cortex*

Data reported in **Chapter 5** were presented at the European Conference on Visual Perception (2015), Liverpool, UK, as:

*Vernon RJW, Gouws AD, Lawrence SJD, Wade AR, Morland AB (2015) The Role of Shape Complexity in the Lateral Occipital Complex*

and at the Yorkshire Vision Network conference (2016), York, UK, as:

*Vernon RJW, Gouws AD, Lawrence SJD, Wade AR, Morland AB (2016) Shape Representations in Lateral Occipital Cortex*

Data reported in **Chapter 5** were accepted for publication as:

*Vernon RJW, Gouws AD, Lawrence SJD, Wade AR, Morland AB (2016) Multivariate Patterns in the Human Object-Processing Pathway Reveal a Shift from Retinotopic to Shape Curvature Representations in Lateral Occipital Areas, LO-1 and LO-2. Journal of Neuroscience 36:5763–5774.*

Data reported in **Chapter 7** were presented at the European Conference on Visual Perception (2016), Barcelona, Spain, as:

*Vernon RJW, Lawrence SJD, Gouws AD, Morland AB (2016) Exploring Curvature Representations in LO-1, LO-2 and Shape-Selective Lateral Occipital Cortex (LO)*

## Chapter 1. Introduction

### *A review on the Role of the Lateral Occipital Cortex in Object Recognition*

#### 1.1 Overview

The central aim of this thesis is to further our understanding of the Lateral Occipital brain regions that appear to largely underpin (human) object recognition. The process of object recognition can appear to be deceptively simple. It is a task that we seem to perform with ease and so it is easy to overlook the underlying processes that drive its successful completion. In this chapter, we will explore these processes to provide a solid theoretical framework for the empirical chapters to follow.

Specifically, we will explore the question of ‘how we recognise that an object *is* an object’; in a sense, approaching ‘high-level’ object processing from a relatively ‘low-level’ perspective. This task involves multiple components; detecting contours, determining boundary ownership, potentially contour completion (e.g. for hidden/occluded objects), figure-ground segmentation etc. These processes all have to act in concert to reliably identify any objects or shapes present in a visual scene. We ask how and where these processes are performed in the brain.

This chapter (and more generally, this thesis) will not address some aspects of the object recognition process, such as recognition in the more traditional sense. We will not be concerned with how specific exemplars of an object are identified (e.g. that mug is *my* mug), or how objects are associated with specific names in the first place (e.g. that object *is* a mug). The key issue to explore is simply how we extract objects from a visual scene; how does our visual system decide that a given set of lines or contours form a coherent, discrete shape? In line with this, we will also generally avoid consideration of top-down effects. It is likely these exist, as certain objects will be hard to identify at first glance, yet a decent guess could be made from inferences. For example, deformable objects such as clothing may be hard to recognise, however a hitherto unrecognisable coat may be ‘identified’ (or its identity guessed at) if hung upon a coat rack. Arguably this is still object recognition, however it is a subtly different process to the one we are primarily interested in – that is, object recognition from a relatively low-level perspective.

There are three core sections in this chapter. The first two sections will briefly introduce our Lateral Occipital regions of interest (ROIs); namely the functionally-defined Lateral Occipital Complex (LOC) and then retinotopically-defined regions LO-1 and LO-2. The third section will then explore properties of the LOC more generally.

## **1.2 The Lateral Occipital Complex (LOC)**

### **1.2.1 Object recognition and the brain**

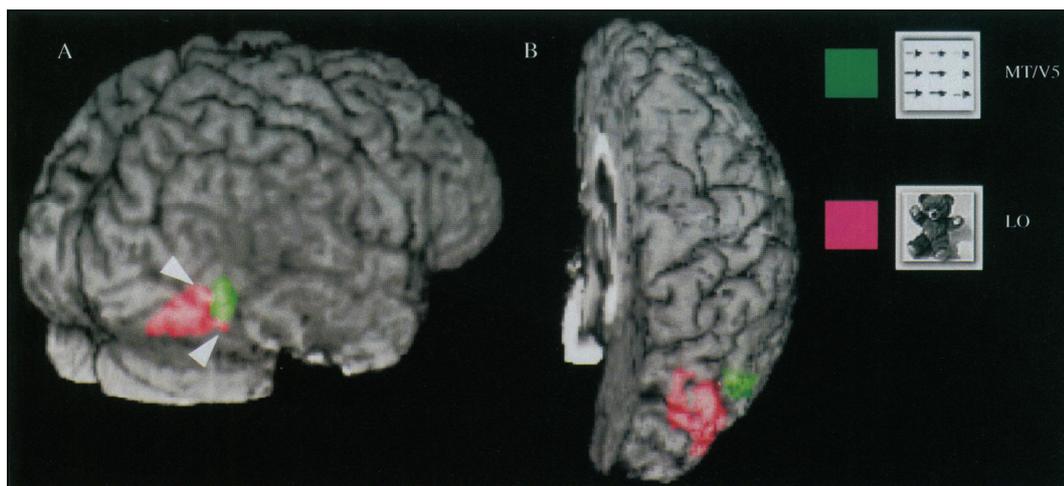
Object recognition is arguably one of the key goals for the visual system. Breaking a visual scene down into discrete components allows us to process and respond to our environment in an efficient manner. It is also a task that likely requires full use of the visual system for reliable performance. For example, we can use a variety of different cues to achieve recognition. We may recognise an apple by its hue, a camouflaged lizard by its motion, or more simply a shape by the luminance contrast between its contours and the surrounding background. Similarly, we can also recognise shapes defined with line drawings, 2D shaded images, photographs etc., suggesting that we can accommodate a wide range of surface cues when attempting to identify a shape. Nevertheless, these varied cues do not give rise to discrete object representations, instead we recognise an apple as an apple whether it is identified from its hue, contours, motion (if falling) etc. This may seem like a trivial observation, but it implies that at some level in the brain there may exist a region (or cluster of regions) that can encode abstracted object representations independently from contextual cues; it is such an area that we are interested in exploring.

Some of the first evidence that such a region may exist came from a seminal paper by Malach et al. (1995). They ran a block-design fMRI study comparing neural responses to pictures of objects versus texture patterns. The objects were achromatic images from one of three categories; famous faces, common objects and 3D abstract (unfamiliar) sculptures. Inclusion of famous faces is less than ideal, as it brings in various confounds linked to face processing (coding of identity, emotional expression etc.). Nevertheless, there were good attempts to explore a variety of different conditions; they also tested both low- and high-pass filtered object images (i.e. removing high and low frequency details from each image respectively), as well

as images under different levels of visual noise (moderate, high, pure noise conditions; each successively reducing the recognisability of the image).

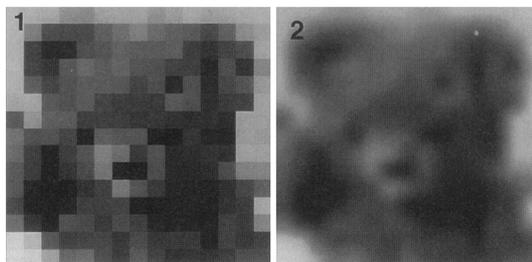
They contrasted these various objects images with phase-scrambled textures, as well as more generic textures such as geometric patterns. Briefly, phase-scrambled textures were produced by performing a 2D Fourier transform on the high-pass filtered objects, converting them into their power and phase spectrum, then adding random noise to the phase spectrum before performing the inverse Fourier transform. Both normal and contrast-matched (i.e. contrast matched to original image) versions were produced. This method attempted to control for visual complexity, producing an unrecognisable image with the same power spectrum as the original object images (but see Chapter 2, Methods, section 2.5.1).

The results of this object-texture contrast identified an inferior, lateral region of the brain (posterior and adjacent to MT/V5) in 15/16 participants (Figure 1.1), which appeared to preferentially respond to objects (or at least the object stimuli used). They termed it the Lateral Occipital Complex (LOC), as they were unable to determine whether they had identified a single homogeneous region, or a cluster of regions, due to the limited resolution of fMRI at the time. They were also generally limited to the study of one hemisphere (the right hemisphere), however in five participants they were able to detect some bilateral activation.



**Figure 1.1.** The Lateral Occipital Complex. The region highlighted in pink reflects the Lateral Occipital Complex (here abbreviated to LO). The green shaded region is motion selective MT/V5. Figure adapted from Malach et al. (1995).

To further characterise this region's response, they next explored its activation to the various manipulated images (frequency filtered/added noise etc.). They found that the LOC responded similarly to both the normal and high-pass filtered object images, but showed considerably diminished activity to both normal and contrast-matched scrambled-face textures. They also demonstrated that the LOC responded similarly to both low- and high-pass filtered faces, indicating that the LOC's response is not simply tied to the spatial frequency of the image. Notably, responses in V1 were broadly consistent regardless of stimulus category, suggesting lower-level visual differences are not driving these effects. Next, they compared familiar objects to the unfamiliar, abstract sculptures, finding both elicited similar levels of activation in the LOC. This indicates that the LOC may not be sensitive to semantic content or familiarity, hinting that it might be a general shape recognition region. They also found evidence that LOC's response depended on recognisability. By increasing the amount of visual noise in an image (in this case famous faces), they showed the LOC's response decreased in proportion to noise. In contrast, V1s response actually increased with noise, corroborating the idea that the LOC is not simply responding to low-level visual features. Furthermore, they also showed that the LOC's responses followed the perceptual experience of the Lincoln illusion, in which a pixelated image becomes more recognisable upon blurring (Figure 1.2.) (Harmon and Julesz, 1973). That is, LOC responses were diminished for a pixelated image, but effectively normal (i.e. equivalent to original, non-pixelated image) when that same image was blurred. Finally, changing the size of the stimuli did not alter responses either, however note that it is not necessarily the same neural populations that are responding to the differently sized stimuli (i.e. they are not claiming size invariance).



**Figure 1.2.** The Lincoln illusion. If you take a pixelated image (1) and blur it (2), we find that it becomes more recognisable. This is known as the Lincoln illusion. Figure adapted from Malach et al. (1995).

In sum, the authors argued that the LOC is a strong candidate for an intermediate link in the object-processing pathway. It is an area that seems relatively untied to low-level visual features, and its activity corresponds to shape regardless of that

shape's familiarity. Furthermore, the fact that the LOC's activity mirrors recognisability hints that it could be linked to our perceptual experiences of shape. The LOC certainly seemed to be a good candidate for a region that could feed into later areas involved in more semantic roles of identifying and categorising shapes.

Subsequent studies replicated these results, demonstrating that bilateral LOC ROIs can be reliably identified across participants using an 'objects versus scrambled objects' functional localiser scan (Grill-Spector et al., 2001). Furthermore, activation elicited by such a localiser is also found to be generally consistent within-participant across different time points (Duncan and Devlin, 2011). This implies that the activated regions likely represent true underlying tuning characteristics of the human brain, rather than spontaneous activation elicited by immediate reactions to the stimuli presented (e.g. such as qualitative evaluation). However, there has been some comment in the literature regarding the limitations of localisers such as these (Friston et al., 2006; but see Saxe et al., 2006). As functional localisers are essential to the empirical work to follow, we will address the potential issues Friston et al. bring up; these will be considered in turn, both from a general perspective and with respect to our own experimental designs. Note that some counterarguments mentioned below were discussed by Friston et al., and so they were not dismissing localisers in their entirety.

### **1.2.1.1 Functional localisers can only show selectivity, not specialisation**

The argument is that a localiser can demonstrate that a certain area is selective for a given stimulus, but it cannot show that the area is the *specific* (or 'specialist') area for that stimulus. In a sense, this is suggesting that by focussing on regions of interest (ROIs), researchers may miss other (potentially relevant) regions activated by their task. This is certainly a valid point, given added relevance by Haxby et al.'s (2001) paper showing that the distributed responses across visual cortex were sufficient to identify stimulus category, even without the maximally responsive areas typically taken as ROIs. Nevertheless, if a given region turns up repeatedly in relatively distinct studies then it seems perfectly valid to want to study the role of that region, even if it is not the only region showing activation. This is arguably the case for LOC. There will likely be multiple regions of activation for object stimuli across the visual cortex, however LOC is reliably activated by a wide array of different object or shape stimuli in almost all participants. As such, understanding

the role of that region is an interesting line of enquiry in its own right. As long as researchers do not focus on their chosen ROI to the exclusion of all other areas, this argument does not seem to undermine the general use of localisers. Indeed, in our own work we will typically analyse not only the Lateral Occipital regions of interest, but also earlier visual areas such as V1-V4. This will allow us to assess functional responses across the visual hierarchy.

### **1.2.1.2 An ROI defined under one context may be inappropriate under different contexts**

This is another valid point; if (for example) LOC did only respond to familiar objects, then it would indeed be a poor choice of ROI in (e.g.) a study considering novel shapes. However, there is mounting evidence that LOC is a region that is responsive to shape regardless of size, position, familiarity, how the shape is defined etc. (Grill-Spector et al., 2001). Therefore, it seems to be a sensible choice of ROI for studies exploring shape or object processing.

### **1.2.1.3 Localisers miss out on potential interactions between the localiser stimuli and main study stimuli**

This argument seems less compelling. The suggestion is that instead of including a separate localiser scan, it could be incorporated into the main study as part of a factorial design. By doing this, it then becomes possible to test for interactions between the two sets of stimuli. Whilst this may be advantageous in certain circumstances, it does come with certain difficulties. First, there is the question of whether there would be any empirical motivation to test the interactions in the first place. Given that localiser stimuli are often quite distinct from the conditions in the main experiment (particularly when exploring lower-level stimuli), this seems relatively unlikely. Introducing the localiser runs into the main experiment would also impinge upon scanning time, reducing overall statistical power; keeping the two scans in separate sessions avoids this issue. Additionally, performing localisers separately helps maintain a clear distinction between the localiser and main study, which might help minimise participant confusion in scenarios when the main study is already quite demanding. This distinction is also important because it is critical to avoid scenarios where the ROI depends in any way on activity tied to the main study (Kriegeskorte et al., 2009). With respect to our own work, we identify ROIs both retinotopically and functionally, and it would not be feasible (or desirable) to include

retinotopy as part of a factorial design. Therefore, to maintain equivalency across retinotopically and functionally defined ROIs, both should be identified in independent scanning sessions. All in all, the reasons to include localisers as part of a factorial design seem limited.

#### **1.2.1.4 Averaging responses over an ROI makes assumptions about homogeneity**

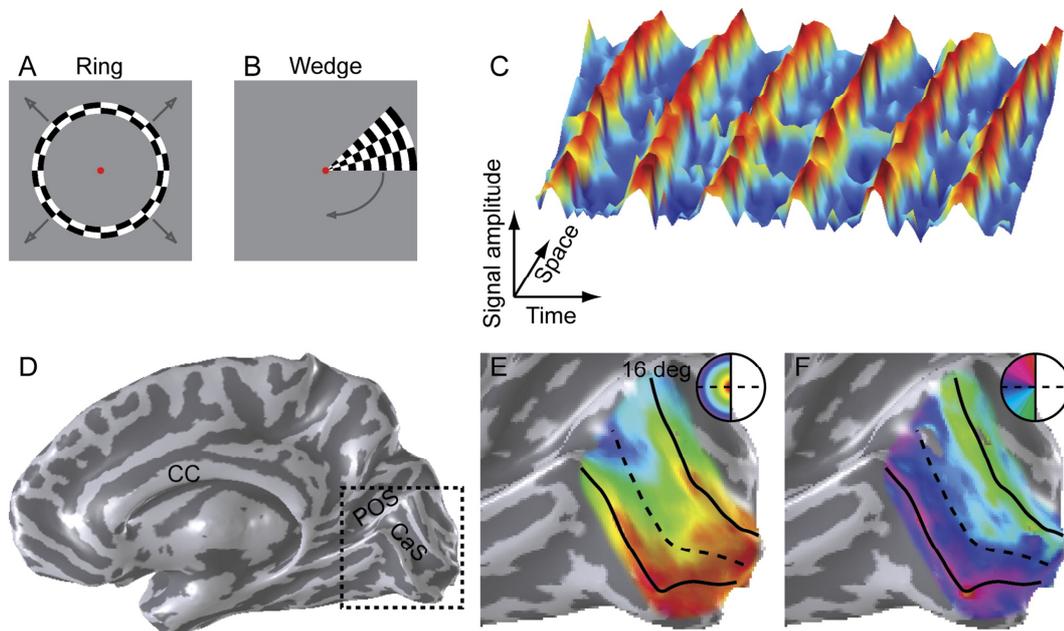
Friston et al. asserted that by averaging responses across an ROI, you assume that the activation will be broadly consistent across the region. This becomes a problem if, for example, only half the ROI was active with the other half suppressed, averaging then would show no signal. However, as Saxe et al. (2006) argue, even a single voxel is still the averaged response of thousands of neurons, so any research design will necessitate this assumption at some level. Furthermore, the use of ROIs does not preclude the additional use of whole brain analysis (with appropriate statistical corrections), which could potentially identify more fine-grained activation within regions of interest. The use of ROIs is perhaps a bit of a blunt tool, but it does minimise the need for stringent thresholding that might otherwise mask genuine effects. There are also more nuanced experimental paradigms that are designed to circumvent this very issue. For example, through the use of adaptation it is possible to infer, from the relative levels of activation, what subpopulations within the defined ROI are actually responding to. The rise of multi-voxel pattern analysis (MVPA) also addresses this issue, as in MVPA we treat an ROI as a cluster of responses rather than a homogeneous whole. Generally, this criticism seems more akin to a cautionary note, and not a reason to avoid the use of localisers in the first place.

In fact, this broadly holds true for all of the points raised so far. Localisers are certainly a valuable tool in modern neuroimaging, as the decades of research on regions such as LOC will attest. The definition and profiling of an ROI through localisers provides a point of convergence for researchers, facilitating collaboration and conversation between what may otherwise be fairly disparate research agendas. However, researchers must also be careful not to overreach, there are certainly limitations with such methods, and careful consideration of these limitations is necessary to ensure findings are discussed in an accurate and scientifically valid manner.

## 1.3 Retinotopy

### 1.3.1 The travelling-wave method

So far we have considered functional localisers, that is, localisers designed to identify regions of the brain that perform specific functions (namely the function of object processing, and corresponding region LOC). However, it is also possible to ‘localise’ regions of the brain by exploiting their organisation with respect to the visual field. It has long been known that there is some kind of retinotopic organisation in early visual cortex (Riddoch, 1917; Holmes, 1918). For example, in visual area V1 (‘striate cortex’) each hemisphere represents the contralateral visual hemifield; as you move anteriorly from the occipital pole, you change from a foveal to peripheral representation, and as you move ventrally to dorsally you change from upper to lower vertical meridian representations. This is termed a retinotopic organisation because the cortical layout of neurons reflects the visual field as projected onto the retina. By using this knowledge, it is possible to design retinotopic mapping procedures (Engel et al., 1994; Sereno et al., 1995; Wandell et al., 2007).



**Figure 1.3.** Retinotopic Mapping. Briefly, through ring (A) and wedge (B) stimuli a travelling wave of activation (C) can be elicited in the Occipital Lobe (D). Through analysis of the phase of that travelling wave, maps of both eccentricity (E) and polar angle (F) can be identified. Figure adapted from Wandell et al. (2007).

Briefly, the general technique is as follows. To map polar angle, participants are presented with a rotating wedge-shaped stimulus, to map eccentricity, participants are presented with expanding ring stimuli. For each voxel, you can then find the best fitting sinusoid that has the same frequency as the cycle rate of your stimulus. In this way, the phase of those sinusoids will represent either the polar angle or eccentricity tuning of that voxel. This approach is termed the travelling-wave method (Figure 1.3), although note that alternative mapping procedures do exist (e.g. population receptive field mapping; Dumoulin and Wandell, 2008, see section 1.3.3). From here, it is possible to use the known retinotopic organisation of visual cortex (Wandell et al., 2007) to demarcate individual visual field maps, primarily by identifying reversals (discontinuities) in the visual field representations. The 'classic' maps start with V1, described earlier, then adjacent to V1 is V2, split into dorsal (V2d) and ventral (V2v) components, each representing one quarterfield. Adjacent to V2 is V3, again split into dorsal (V3d) and ventral (V3v, sometimes VP) components. By continuing this process of identifying reversals, it is possible to classify further areas including V3A/B, V4 etc.

This is a much more principled method of localising ROIs as it is largely removed from any assumptions about the functional properties of those areas, and crucially the retinotopic stimuli used will generally be distinct from the main experimental stimuli. This avoids potential criticisms of functional localisers, for example it could be argued that an object localiser biases main studies incorporating objects towards significance (as if you identify an area that responds to objects, then any study using stimuli where one condition is more object-like than the other is potentially biased towards showing significant differences).

Of course, this method only works if the area of interest is retinotopic in the first place; a condition that is not necessarily true for extrastriate visual areas. Indeed, if visual representations are to become more abstract as you move away from V1 into extrastriate cortex (so that they can detect features or shapes regardless of retinal position; i.e. location invariance), then it may actually be a necessity for these areas to be non-retinotopic.

### 1.3.2 Early visual cortex

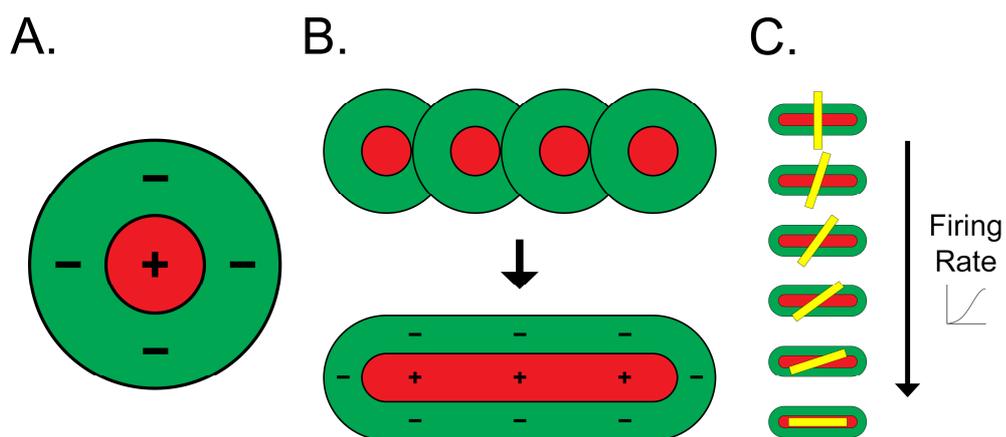
Whilst this thesis will predominantly focus on Lateral Occipital cortex, we will briefly discuss early visual cortex as it allows discussion on the concept of low- versus high-level visual processing. For the current purposes, we will consider the retinotopically defined regions V1 through to V3 as early visual cortex, with greater emphasis on V1 as the roles of V2 and V3 are much less understood (Peirce, 2015).

Vision is ultimately an integrative process, a combination of two influences. We take the *physical input* of light falling on the retina, and, combined with *past experiences and visual heuristics*, we construct a coherent percept from it. This can be readily demonstrated, for example, take the letter *f*. In terms of physical input, the contrast of black against white (or low versus high luminance) allows us to detect the lines comprising its form. Interpreting those lines could be ambiguous; they could be perceived either as a whole, as four segments, two abutting vertices etc. However, the heuristic of ‘good continuation’ (Wertheimer, 1923; Field et al., 1993) resolves that ambiguity and we perceive one curved line overlaid with one straight line.

The distinction between low- and high-level processing reflects the relative strengths of (or balance between) these two influences. Low-level vision will be more concerned with the ‘raw’ processing of basic visual cues; brightness, contrast, orientation, colour, depth, motion, etc. Areas we talk of as ‘high-level’ will integrate across these various cues, using heuristics to identify common patterns, and thus coherency. However, we should stress that the low-/high-level distinction is not binary; it is likely all visual areas will integrate to certain extents. We have already discussed to some degree how an object may be recognised based upon any of the cues listed above, and so we may consider object processing ‘high-level’. Given that high-level processing must ultimately be based upon low-level input, we aim to control for low-level cues as tightly as possible in visual experiments. This should concentrate stimulus differences in the high-level features of interest, rather than confounding low-level influences. To control for such influences, we must first understand how they are processed.

To explore the processing of low-level visual inputs, we will consider a single cue; that of orientation. First, after phototransduction (detection and response to light) signals from the retina travel (predominantly) along the retino-geniculate pathway,

passing through the lateral geniculate nucleus (LGN) and into V1. Cells in both the retina (ganglion cells) and the LGN contain ‘centre-surround’ receptive fields (Figure 1.4A). The size and ratio between centre/surround will vary based upon multiple considerations (e.g. foveal versus peripheral location, current light levels etc.), but the fundamental principles of their response profile remains the same. An on-centre, off-surround cell will be stimulated whenever a) light increases in the centre or b) light decreases in the surround. For an off-centre, on-surround cell, the situation is naturally reversed. The combination of these two receptive field types allow rapid detection of light intensity changes (whether incremental or decremental). Critically, the fundamental processing step here is that of comparison. Gross light levels are of relatively little importance, indeed there are mechanisms such as adaptation/gain control that serve to discount absolute light levels. Instead, we are highly sensitive to changes across space (contrast) and changes across time (onset or offset of stimulation).



**Figure 1.4.** Receptive fields in early visual cortex. (A) An on-centre, off-surround cell as typically found in retinal ganglion or lateral geniculate nucleus cells. (B) The combination of multiple centre/surround cells allows tunings for more complicated stimuli such as oriented lines. (C) A consequence of this response pooling is that responses are not binary; they instead show Gaussian tuning with the peak firing rate centred at the optimal stimulus type (here, a horizontally-oriented line).

Simple cells in V1 will sample across multiple of these centre-surround receptive fields, allowing for responses that are tuned to more complicated stimulus properties (Hubel and Wiesel, 1959). An example is the orientation detection cell shown in

Figure 1.4B. This cell will show peak responses to horizontally-oriented stimuli, especially when flanked by a surround of lower luminance. In essence, it is a line detector. These simple cells are distributed throughout V1 (in a markedly systematic fashion) to sample across the full range of orientations and visual field locations available. Furthermore, centre/surround cells can be pooled in different ways, and the simple cells themselves can also be pooled (via complex cells) allowing responses to increasingly complex stimulus properties (edges, corners, curvature etc.). Again, this illustrates the inherently constructive nature of visual processing. Representations are built by integrating and abstracting across those responses that came earlier in the visual system (for simplicity, we are not considering top-down effects here).

A final consideration is that for the hypothetical cell shown in Figure 1.4B, a horizontal line will lie over the *inhibitory* regions of some of the centre/surround cells that comprise it. It will therefore not elicit maximal activation in those centre/surround cells. However, the *net* response across all cells sampled will nevertheless be positive. Pooling responses in this fashion consequently means that a horizontally-tuned cell will not respond in a binary manner, only firing in response to the perfect stimulus. Instead it will show Gaussian tuning, with increasing firing rates as a given stimulus tends towards its target orientation (Figure 1.4C). Essentially, this means that there is inherent uncertainty in neuronal responses, and the visual system must make a probabilistic ‘best guess’ to build the visual scene. This is likely where the previously discussed heuristics come into play, as they will allow better predictions based upon prior experience and known properties of the visual world.

The purpose of this brief review of early visual processing is to highlight three critical aspects of neuronal tunings. First, we note that the visual system is inherently *comparative*. This implies that to match stimuli in terms of low-level features, we need to control for properties such as contrast and orientation energy both across space and across time (i.e. across successive stimuli). Simple considerations such as ensuring stimuli take up approximately equal spatial extents, or equalising image histograms (thus mitigating sudden luminance changes) can address this to a certain degree. Secondly, the visual system is a *constructive* process. This means that an object recognition system will not spontaneously appear, it will be built upon

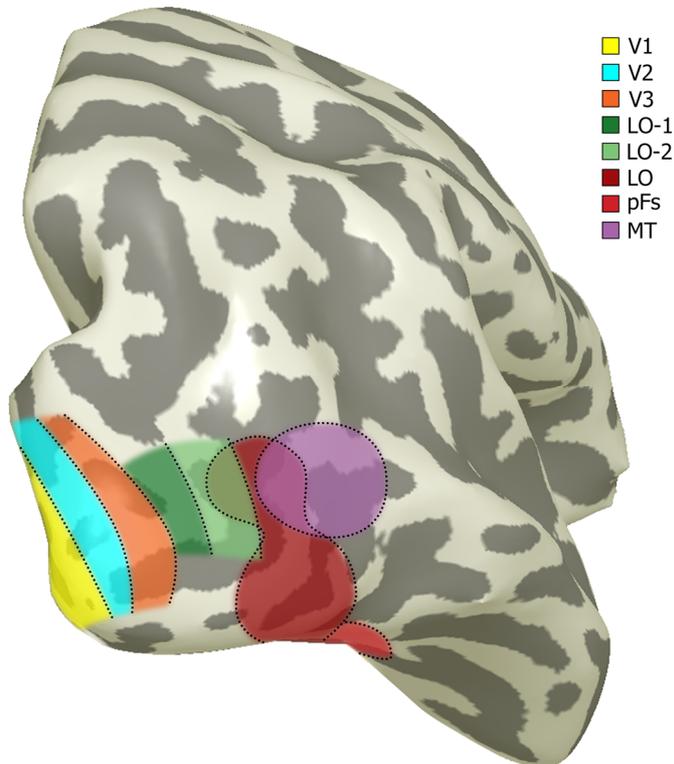
foundations based on the low-level detection mechanisms in early visual cortex. On this basis, there should be intermediate representations between low-level features such as those discussed above, and more ‘abstract’ high-level dimensionality. Probing such intermediate representations could yield valuable insights into the later mechanisms that underpin object recognition.

The third and final consideration is that neurons are *tuned*, not selective, to their preferred stimulus type. This again has interesting consequences for object recognition, and suggests a means to actually probe neural representations across visual cortex. It is immediately obvious how low-level features can systematically vary, for example orientations span 0-360°. However, as such features become increasingly integrated and abstracted, the nature of the tuning becomes less apparent. Nevertheless, it is likely that objects are still encoded in this probabilistic fashion, therefore a critical question to ask is which stimuli elicit more *similar* patterns of activation in high-level visual areas? Take a cell responding to a 90° bar for example. If we found that it responded similarly to a 70° bar but not a 45° bar, then we have good evidence that it is orientation selective. Instead, if that cell responded to all orientations but not a square, then perhaps orientation is incidental to its tuning and it is instead tuned to some other stimulus property (e.g. aspect ratio). Similarly, if we can determine which shapes elicit either similar or dissimilar patterns in high-level cortex, we can attempt to explore underlying neuronal tunings by considering the commonalities across those stimuli. Indeed, such an approach (often referred to as representational similarity analysis; Kriegeskorte et al., 2008) is becoming increasingly popular in the literature. In sum, we hope that this section demonstrates the importance of understanding low-level visual processing, to inform design choices when probing high-level visual cortex.

### 1.3.3 LO-1 and LO-2

The primary reason for introducing retinotopy was to discuss two visual field maps that have been identified relatively recently; LO-1 and LO-2 (Larsson and Heeger, 2006). As their names suggest, these maps lie in Lateral Occipital Cortex and they potentially overlap with the LOC; the region defined by object localisers. Relative to other visual areas, LO-1 lies adjacent to V3d, containing a full hemifield spanning the lower-to-upper vertical meridians. LO-2 then lies adjacent to LO-1 (extending to or nearly to V5/MT+), also containing a full hemifield with a representation

mirroring that of LO-1s. A rough schematic can be seen in Figure 1.5. Both areas share a foveal representation with V1-V3.



**Figure 1.5.** Region of interest overview. Here, we show the posterior end of the lateral surface of the right hemisphere of one representative individual. The approximate locations of retinotopically defined regions V1-V3, LO-1 and LO-2 (plus MT), as well as functionally defined regions LO and pFs are outlined. ROI locations have been taken from data to be shown later (Figure 2.8; Chapter 2: Methods, section 2.6.5). Note that the two roughly circular sections defined as LO have sometimes been split into LOa and LOb ROIs (e.g. Vinberg and Grill-Spector, 2008). The approximate position of MT was estimated based upon retinotopic ROIs TO-1 and TO-2 (*not shown*); regions that are known to possess some overlap with MT (Amano et al., 2009).

Critically, Larsson & Heeger found that both LO-1 and LO-2 responded significantly more to objects over scrambled objects, with LO-2 showing the (significantly) greater modulation. However, instead of using a conventional contrast, object activation was identified by fitting sinusoids to the data and assessing the phase, which corresponded to the stimulus that evoked the largest response (objects or scrambled). This analysis appeared to find large swathes of visual cortex preferring objects over scrambled, suggesting that whilst LO-1 and LO-2 may be object sensitive, they do not necessarily overlap with LOC as it is traditionally defined.

However, later papers have since provided further evidence for this overlap (Sayres and Grill-Spector, 2008; Amano et al., 2009; Silson et al., 2016), corroborating their findings.

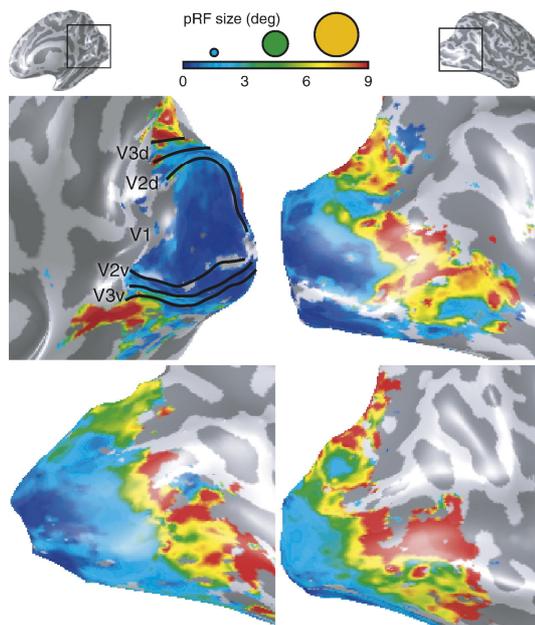
#### **1.3.4 Retinotopic organisation of the Lateral Occipital Cortex**

The overlap between LO-1/LO-2 and LOC merits further consideration as earlier studies suggested LOC contained little to no retinotopic organisation (Levy et al., 2001; Tootell, 2001). If LOC does overlap with retinotopic regions, then these would obviously be contradictory findings. A potential source for this discrepancy comes from Dumoulin and Wandell (2008), who introduced a new approach for conducting retinotopy termed the population receptive field method (pRF).

The underlying principles of pRF are comparable to that of retinotopy. It can still employ rotating wedges and contracting or expanding rings, albeit with the introduction of blank periods to replace certain visual field locations (i.e. instead of the wedge transitioning from say  $60^\circ$  to  $45^\circ$ , it may transition from  $60^\circ$  to a blank (mid. grey) screen before then continuing as normal after a short interval). In addition to these standard stimuli, pRF can also incorporate bars that move across the screen at various orientations. From the responses to these stimuli, it is then possible to find for every voxel the best fitting Gaussian model which includes both (x, y) co-ordinates for the visual location eliciting peak activity, and  $\sigma$ , a measure of spread. It is this spread that is of crucial interest here, as it acts as a proxy for receptive field size; the area of the visual field over which a given voxel (or specifically, neurons within that voxel) will respond.

Using this method, Dumoulin and Wandell found substantial differences in the response properties of Lateral Occipital Cortex and V1. Firstly, V1 showed much greater modulation when compared to Lateral Occipital Cortex, suggesting narrower receptive fields. Corroborating this, blank sections only diminished a given region of V1's BOLD response when that blank section replaced a retinal position the region was tuned to. In contrast, a blank section always dropped the Lateral Occipital BOLD response. This suggests that even if regions within Lateral Occipital Cortex had preferred visual locations, those regions still broadly respond to stimuli anywhere in the visual field (or at least anywhere in the region covered by the pRF mapping stimuli). They also generally found less lateralisation in Lateral Occipital

Cortex than in V1, again suggesting that Lateral Occipital regions are much more broadly tuned to visual field location. Quantifying this, they estimated that Lateral Occipital receptive fields are approximately five times larger than that of V1's (Figure 1.6). The key point of this finding is that earlier studies, with coarser resolution fMRI, are unlikely to have been able to identify the more subtle changes in BOLD response that are found with large receptive fields; hence the earlier (likely erroneous) conclusions that there is no retinotopic organisation in Lateral Occipital Cortex.



**Figure 1.6.** Population receptive field (pRF) sizes. The pRF size for a given region provides an estimate for the area of visual space over which that region will respond. pRF sizes (shown here in degrees of visual angle) differ considerably across the Occipital lobe, with pRF sizes in V1 being considerably smaller than those in later Lateral Occipital regions. Figure adapted from Dumoulin & Wandell (2008).

Later studies have since directly compared LOC (defined using standard contrasts) to LO-1 and LO-2 (Sayres and Grill-Spector, 2008; Amano et al., 2009; Silson et al., 2016), identifying overlap between the regions. Sayres and Grill-Spector for example found an average 8.1% ( $\pm 10\%$ ) overlap between LOC and LO-1, and 18.8% ( $\pm 11.8\%$ ) overlap with LO-2. Typically, it was the more dorsal, posterior component of LOC that overlapped these regions, with LOC continuing anteriorly/ventrally below V5/MT+. Furthermore, they found LOC still had some retinotopic organisation, even when excluding the LO-1/LO-2 overlap. It had a strong foveal bias, a slight contralateral bias (also found by Hemond et al., 2007), and intriguingly a consistent bias towards the lower vertical meridian (perhaps as objects are typically located in the lower field of view). When conducting a split-half analysis (correlating responses across odd and even runs), Sayres & Grill-Spector showed that objects in similar retinal positions elicited very similar activity between runs, indicating that

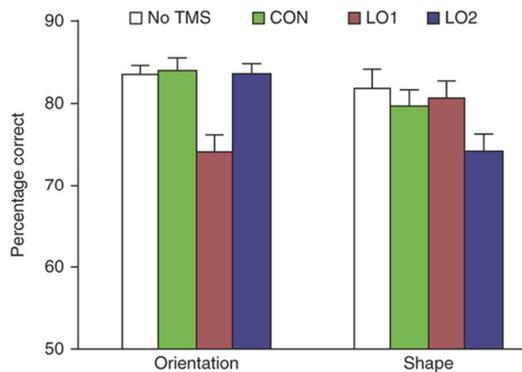
LOC's activity is at least partly sensitive to position information. They also identified weaker correlations when comparing across categories (faces, animals, cars & sculptures), which appeared to be independent of the position effect. That is, the evoked response to (e.g.) a car in one run will be more similar to car responses in another run, compared to any other category. However, this does not necessarily indicate category selectivity; it could simply be low-level similarities driving the correlation (e.g. cars will generally have similar aspect ratios and tend to contain curves in similar places, which are unlikely to be present in alternative categories).

### **1.3.5 Additional properties of LO-1 and LO-2**

If LO-1 and LO-2 are part of object-selective LOC, then it raises the question of what role they play. As previously discussed, both areas are object sensitive, but LO-1 (not LO-2) also appears to have some orientation selectivity (Larsson et al., 2006), although this may be relatively weak (Sapountzis et al., 2010). There is also evidence that these areas are sensitive to illusory contours (defined by vertical/horizontal misalignments between two sets of orientated lines), with LO-1 and LO-2 both displaying adaptation when the same illusory contours were repeated (Montaser-Kouhsari et al., 2007). LO-1 and LO-2 do not appear to encode colour (Brouwer and Heeger, 2009), however Larsson and Heeger (2006) found that whilst neither area preferred moving over stationary dots, they did respond to motion boundaries (see also Larsson et al., 2010). As such, it is possible these areas will still respond to colour information if it used to define shapes or contours.

In their initial study, Larsson & Heeger hypothesised that LO-1 may be extracting boundary information, and LO-2 representing shape, perhaps then passing this information forwards to the more anterior/ventral regions of LOC for location-invariant processing. The most compelling evidence for such a division in function comes from a transcranial magnetic stimulation (TMS) study, which in fact identified a double dissociation between these two areas (Silson et al., 2013). The evidence considered so far has suggested that LO-1 may play some role in orientation discrimination, whereas LO-2 appears to be the more object-sensitive region. Based upon this, Silson et al. designed stimuli that incorporated these distinct visual features, having participants perform two tasks. Firstly, orientation discrimination; participants judged whether a test stimulus (circular aperture filled with a sinusoidal black & white grating) was more horizontal or vertical than a

reference stimulus. Secondly, shape discrimination; participants judged whether a test three-lobed radial frequency pattern (essentially a clover-shaped outline) was spikier or smoother than a reference shape (i.e. curvature judgements). Performance (accuracy) for both of these tasks was matched across participants through the use of psychophysics. These two tasks were then performed during TMS, and it was found that TMS to LO-1 *solely* disrupted the orientation task, whereas TMS to LO-2 *solely* disrupted the shape task (Figure 1.7).



**Figure 1.7.** LO-1/LO-2 double dissociation. TMS to LO-1 impairs orientation processing, whereas TMS to LO-2 impairs shape processing. No TMS and TMS to a control site both failed to impair performance on either task. Error bars reflect SEM. Figure adapted from Silson et al. (2013).

This is an intriguing finding in many ways. Firstly, despite these two areas being adjacent to one another, they appear to maintain at least a degree of functional independence. This would suggest parallel pathways are needed to feed to these two areas (as noted by Silson et al.). Secondly, there is the question of which processes were actually being disrupted. Neurons in V1 could arguably have ‘solved’ the orientation task, as local edge information (and small receptive fields) should be sufficient to do so. Instead, at least part of the role appears to be falling to LO-1, which has much larger receptive fields (Amano et al., 2009). It is possible that LO-1’s role is actually to pool orientation cues across larger regions of visual cortex. This may help determine the global orientation of objects in visual space, as whilst a given object might have myriad local orientation cues, overall such cues will likely fall parallel to that object’s principal axes (imagine e.g. orientation cues on a tree’s bark). If this were the case, then the fact that local cues were sufficient to solve the task would merely be coincidental. This would also place LO-1 into a slightly different role to the one Larsson and Heeger initially hypothesised. Nevertheless, their hypothesis for LO-2 could still stand, but it is possible that its role may extend to the boundary analysis initially placed upon LO-1, explaining why LO-2 appears causally involved in the curvature/shape processing task from Silson et al. In such a

scenario, LO-1 may be passing pooled orientation cues through to LO, and LO-2 passing general shape profiles. As these tasks appear to be independently governed, such a system would likely have some redundancy. However, this may actually be beneficial; LOC could extract and process objects based upon mutual, complementary sources of information; in scenarios where one source fails, the other may prevail.

To conclude this section, LO-1 and LO-2 have been identified as two retinotopic regions that appear to be playing discrete roles in the visual system. These roles do not necessarily tie in to object recognition, however given their object sensitivity and overlap with LOC, this certainly seems plausible. There is relatively little literature on these two regions, perhaps due to the inherent difficulty in their identification. A central aim of this thesis will therefore be to address this gap; exploring both the functional properties of these two regions, and how they relate to the later LOC.

## 1.4 Properties of LOC

### 1.4.1 Section overview

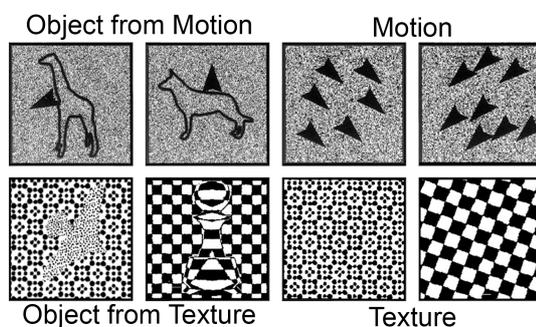
In the final section of this chapter, we will return to the LOC and explore the numerous papers that have studied its properties. The section will be broken down into three main topics:

- 1.4.2. **Cue invariance:** which visual (colour, motion, luminance, etc.) and physical (size, location, orientation, etc.) cues will elicit activity in LOC. Furthermore, if cue invariance is found, then is this property represented in individual neurons, or by the distributed responses across LOC?
- 1.4.3. **Boundaries and contours:** is LOC simply responding to the outer contours of a shape, or does its responses require a more complete object representation (e.g. inclusion of inner shading)?
- 1.4.4. **Perceptual experience of shape:** lower-level visual areas represent the visual field in various ways, but these do not necessarily correspond directly to our perceptual experiences. Is the same true of the LOC; does it encode objects in some abstract form, or do its representations reflect our everyday experiences of object perception?

Note that for consistency, in the following sections ‘LOC’ (Lateral Occipital Complex) will refer to the area(s) identified using the standard objects versus scrambled objects contrast (even when the original study refers to this area as LO). The posterior, dorsal section of LOC (adjacent to/overlapping with LO-2) will be termed LO, and the more anterior, ventral region is referred to as pFs (for posterior fusiform gyrus).

### 1.4.2 Cue Invariance

As previously discussed, we do not rely on a single cue to identify objects, we incorporate a range of different visual properties to aid object recognition. For example, in addition to simple luminance boundaries, we can identify objects defined using texture, motion, colour etc. Indeed, it has been hypothesised that trichromacy (having three cones instead of two, allowing us to discriminate reds from greens) may have evolved in part to identify ripe fruit by its colour contrast with the surrounding foliage (Regan et al., 2001), highlighting the importance of alternative cues for object processing. This raises the question of whether LOC is the region that responds to these disparate cues, or if they are all processed independently before converging later in the visual system (if at all).



**Figure 1.8.** Objects from motion and texture. Objects can be defined from various cues, included motion (top row) and texture (bottom row). The dots in motion conditions move in the direction of the arrows, object outlines are for illustrative purposes only. Figure adapted from Grill-Spector et al. (1998a).

Evidence that LOC does in fact respond to these varied cues came soon after the regions discovery. Grill-Spector et al. (1998a) localised regions that responded preferentially to *motion-defined* objects (coherent motion within a shape set against a static or contrasting-motion background) over full-field coherent motion, as well as *texture-defined* objects (objects defined either with a different texture to the background, or with the same texture distorted) over full-field textures (Figure 1.8). Both of these ROIs overlapped considerably with LOC (75%+ overlap) and with each other. Whilst the texture-defined objects could potentially be identified using

luminance contrast around the outer boundary, this is unlikely to be the most salient cue, and luminance contrast could not have identified the motion-defined objects. As such, it not only provides evidence that LOC responds to varied cues, but also that LOC can respond to second-order stimuli. However, due to the limited spatial resolution of fMRI, this study could not demonstrate that LOC is a homogenous, cue invariant region. Cue invariant neurons are within the realm of possibility (Quiroga et al., 2005), but an alternative remains; LOC could instead contains clusters of neurons each selective for different cues.

One approach that might circumvent the spatial resolution limitation is the use of adaptation (e.g. Grill-Spector et al., 1999). In general, say you have a region that responds similarly to Stimulus 1 (S1), and Stimulus 2 (S2); you can present S1 for a relatively long duration of time (causing adaptation), and then present S2. If the BOLD response to S2 is diminished after the adaptation period, then it suggests that there is at least some degree of overlap in the neurons coding S1 and S2 (i.e. there will be some neurons that are sensitive to *both* stimuli). In contrast, if the response to S2 is not diminished, then it suggests that these two stimuli are encoded by discrete sets of neurons. There is substantial evidence demonstrating that LOC can adapt to object stimuli in this way (e.g. Vuilleumier et al., 2002), indicating that this is a viable approach to examine cue invariance.

Grill-Spector et al. (1999) used this method to show that LOC can adapt to objects with different sizes, and in different positions, although the process of adaptation was slower compared to identical stimuli. No such adaptation was found when different objects were presented repeatedly, even if those objects were from the same semantic category (e.g. dogs). However, note that generally only partial adaptation was observed; activity was not reduced to the level observed if identical objects were presented repeatedly. This study also noted a slight dissociation between the more dorsal/posterior section of LOC (LO) and a more ventral/anterior region termed pFs (posterior fusiform gyrus), which also preferred objects over the scrambled counterparts. It was this latter region that showed most adaptation, perhaps suggesting a trend towards decreasing specificity as you move away from lower-level visual areas.

This finding of slight tolerance to size manipulations (with increasing tolerance as you move anteriorly) has been replicated (Sawamura et al., 2005), though a study using multivariate pattern recognition found slightly contradictory results (Eger et al., 2008). Eger et al. managed to predict observed category (kettle/chair) with accuracy that was significantly greater than chance from the distributed response profile of voxels in LOC. Critically, this was true even when the objects were at different sizes or viewpoints across sessions, indicating some level of invariance in the profile of responses. However, in this study it was the more posterior section of LOC (namely, LO) that allowed better predictions, not the anterior regions (pFs) as previously found. The authors proposed a number of explanations, for example pFs may have lower signal-to-noise (responses were generally weaker in this area and prediction accuracy was close to chance), or perhaps there could be generally stronger adaptation in pFs, compared to LO. This is plausible, as pFs shows stronger adaptation than LO even for repeated identical images (Grill-Spector et al., 1999), and it could cause diminished sensitivity when looking for invariance as responses may saturate. Nevertheless, the key point here is that neurons in LOC appear to possess at least some invariance to changes in size, and perhaps viewpoint. It is also worth noting that the ability to predict category does not necessarily imply *category* selectivity (in a semantic sense); it seems more likely that the differential responses found in LOC reflect the general differences in the shape profiles that these two categories have.

Another question to consider is whether LOC simply responds to these varied cues independently, or if it can combine them to facilitate processing. Self & Zeki (2005) addressed this using both motion and colour defined shapes. Specifically, the shapes were defined with a region of coherent motion or colour, against a background of random motion or colour respectively (all stimuli were comprised of dots). Objects were also defined using both cues simultaneously to test potential cue integration (note that both cues were present in all stimuli, it was only coherence that changed). The level of coherence within each shape (e.g. the proportion of dots moving in the same direction) was adjusted to match ease of detection across participants and cue type.

Self & Zeki first aimed to find regions that preferentially responded to 100% coherent motion- or colour-defined shapes over stimuli with no coherence

(essentially a localiser). This identified an area consistent with previous reports of LOC (plus regions in V3A, V5 & superior parietal lobe), broadly replicating the earlier evidence for cue insensitivity in LOC (Grill-Spector et al., 1998a). When searching for regions that preferred both colour and motion cues together over either cue alone, only a sub-region of (right) LOC emerged as significant. A final study demonstrated that the region just identified would adapt strongly to shapes regardless of the cue (motion/colour) used to define them (i.e. it adapted to similar extents if cues alternated, or the same cue was used repeatedly). In contrast, alternative regions of LOC would only adapt when their preferred cue was used, indicating that this is not just a general property of adaptation. Together, these findings provide compelling evidence that at least some sections of LOC may integrate disparate cues, likely improving the reliability of object detection.

In sum, the findings discussed indicate that LOC is relatively insensitive to changes in the visual cues used to define an object (colour/motion etc.), and perhaps to a lesser extent the physical properties of that object (its size/position etc.).

Furthermore, there is some evidence suggesting that this insensitivity to extrinsic stimulus properties generally increases as you move anteriorly, away from LO and towards pFs. There is also modest evidence that the invariance is in part due to cue-insensitive neurons, although the alternative hypothesis of tightly clustered cue-specific neurons remains possible. Either way, the findings lend support to Malach et al.'s (1995) proposal that LOC is an intermediate area in the object-processing pathway. It could play a vital role in identifying objects regardless of context, before passing them on for more explicit identification. This idea also fits with LOC's insensitivity to semantic information (e.g. Grill-Spector et al., 1999; Kim et al., 2009), as the process of categorising or naming objects would be expected to take place after initial shape segregation.

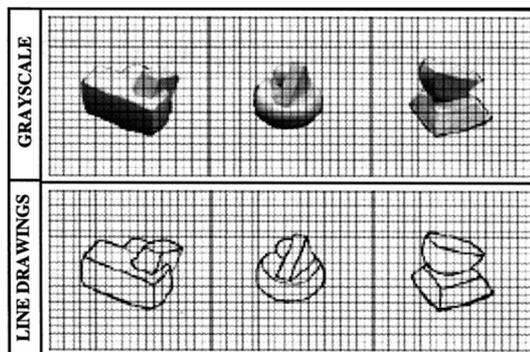
### **1.4.3 Boundaries and Contours**

It is notable that all the manipulations discussed in the previous section would only affect the boundaries of an object. Size or viewpoint manipulations just scale or translate the boundary, and varying the cue that defines the object (motion, colour etc.) just changes the type of contrast that highlights the object's contours. Given that LOC appears to respond regardless of these various manipulations, it appears plausible that contour processing is LOC's primary function. Of course, real objects

contain many more (potentially highly relevant) details, for example it is rarely possible to get depth information from boundaries alone; we need to take into account additional properties such as shading or perhaps internal contours. This section will explore LOC's response to boundaries; just how important are they, and do any other details matter?

### 1.4.3.1 Boundary Importance

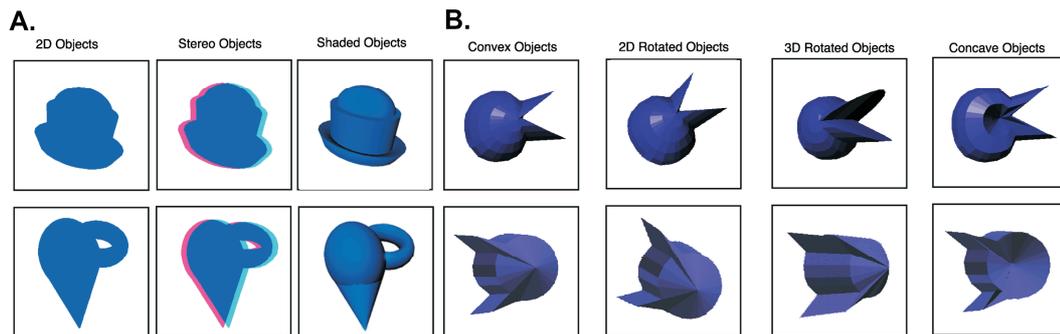
The importance of inner details such as shading was addressed directly by Kourtzi & Kanwisher (2000) in three stages. First, they compared the activity produced by greyscale photographs of novel 3D objects to the activity produced by line drawings, which were created by tracing those same objects (including the internal contours, which provided some depth information) (Figure 1.9).



**Figure 1.9.** Greyscale and line drawing objects. Line drawing stimuli were generated by tracing both the outer and inner contours of greyscale objects. Figure adapted from Kourtzi & Kanwisher (2000).

Both conditions were contrasted with scrambled alternatives, and half the data was used to define LOC (objects in either form over scrambled counterparts), the other half to explore responses in that region (ensuring localiser independence from the main analysis). The results suggested that LOC responded as strongly to both stimuli types; no significant effect was found for shape cue (greyscale '3D' object versus line drawing). This was true even when the LOC was defined with greyscale > scrambled only, and broadly true for line drawings > scrambled only (although those voxels also responded to greyscale scrambled images). Furthermore, a second experiment using an event-related adaptation design demonstrated that LOC would adapt to similar extents regardless of whether objects were repeated in the same or different cues. This was not true if different objects were repeated in the same cue type. This suggests that the adaptation is specific to stimulus shape, and consequently that stimulus shape has been 'abstracted' from cue type, implying both cues are activating the same populations of neurons. A final experiment found no significant differences between LOC's activation for line drawings, 2D outlines (no

depth information) and occluded outlines. Together, these results strongly suggest that even if LOC is sensitive to 3D cues (whether from lines or shading), it does not preferentially respond to them. However, it should be noted that whilst not significant ( $p > 0.3$ ), greyscale objects did elicit greater activity in LOC than line drawings (albeit with more variance). As such, it remains possible that the current study simply lacked the power or sensitivity to find reliable differences between the two cue types.



**Figure 1.10.** 2D versus 3D stimuli. In two experiments, Kourtzi et al. (2003) compared either (A) silhouettes to 3D shaded objects (plus stereo objects based upon red-green anaglyphs) or (B) the use of 3D cues (convexity versus concavity) to distinguish between object exemplars. Figure adapted from Kourtzi et al. (2003).

LOC may not prefer 3D shapes, but there is evidence to suggest that it is sensitive to the cues that define them. Kourtzi, Erb, Grodd, & Bühlhoff (2003) explored the response properties of LOC by first contrasting images of 3D shaded objects with silhouettes (Figure 1.10A), and then by using 3D shading to dissociate objects with the same outer contours (by manipulating concavity and convexity) (Figure 1.10B). Their first study actually found release from adaptation when switching between 3D shaded objects and silhouettes of those same objects, which seems to contradict the previous findings from Kourtzi & Kanwisher above. However, there are (at least) two potential explanations for this discrepancy. The previous study's line drawings contained inner details that provide depth cues, whereas silhouettes only contain outer contours. As such, it is possible that the ambiguity inherent in silhouettes causes them to appear as distinct shapes when compared to their 3D shaded counterparts. Alternatively, 3D shading can separate an object into parts (e.g. the brim versus the crown of a hat in Figure 1.10A), and LOC might be responding to these parts independently. Due to their inner contours, the line drawings in Kourtzi & Kanwisher's study could also have been broken into parts, whereas silhouettes can

only be perceived as a whole. Both possibilities would explain why line drawings and 3D shaded images would be treated as equivalent in one study, but silhouettes and 3D objects would be treated as distinct in another. Note that general depth information is not sufficient to cue differential responses in LOC. Kourtzi et al. also compared stereoscopically defined silhouettes (which appeared on a different depth plane) to their 2D counterparts, finding adaptation to one diminished responses to the other. This reinforces the idea that depth information only becomes a relevant cue for LOC when it serves to dissociate one object from another. When considered alongside signs of size invariance earlier (Grill-Spector et al., 1999; Sawamura et al., 2005; Eger et al., 2008), it also indicates that LOC may possess some invariance to spatial location, as it could accommodate the changes in size and depth that would otherwise differentiate near from far objects.

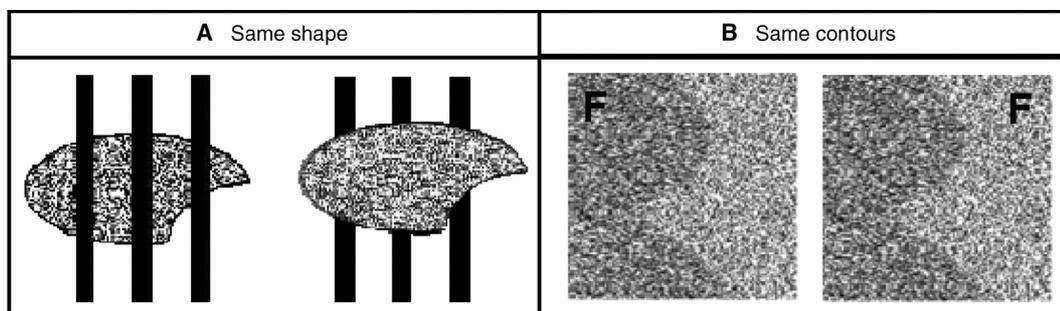
In the second component of Kourtzi et al.'s (2003) study, shading was used to give objects (that had the same external contours) either a concave or convex appearance. This manipulation released LOC from adaptation, corroborating the idea that LOC is sensitive to these depth cues and is not responding to outer boundaries alone. Intriguingly, adaptation *was* observed when objects with the same shading (whether concave or convex) underwent small (30°) rotations (both in frontal plane, and in depth). This indicates that outer contours may not be the most salient feature for LO, with perhaps perceptual similarity being the primary cue. However, later studies have found evidence for rotation sensitivity in these regions (Andresen et al., 2009), so it would be premature to draw general conclusions from these data alone. Finally, when comparing responses between the more posterior parts of LOC to pFs, it was actually found to be pFs that was the region most sensitive to changes in depth. This seems to contrast with earlier studies suggesting an anterior gradient of increasing invariance. However, as noted earlier Eger et al. (2008) also found that it was the posterior regions showing most invariance. The source of this discrepancy is unclear. It is possible that underlying methodological differences are driving the differences, alternatively adaptation may just be inherently variable (perhaps depending on uncontrollable influences such as the attentional state of the participant). Nevertheless, the aim of this section was to explore LOC's sensitivity to shading and depth cues, and in this matter the evidence is in broad agreement. LOC does display

sensitivity to 3D structure, and it can use this information to distinguish between shapes that would appear identical based upon outer contours alone.

We have now shown evidence suggesting that LOC will respond to shapes defined solely by their boundaries, but also that these boundaries can be overridden when additional information is present. This raises the question of how flexible LOC's boundary representations actually are.

### 1.4.3.2 Boundary Flexibility

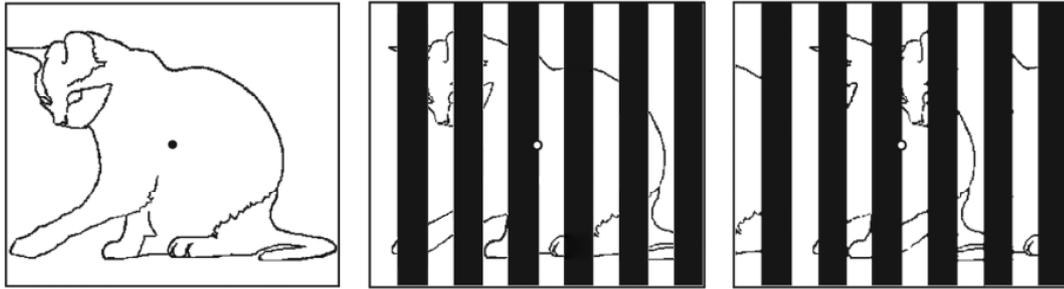
There is substantial evidence that LOC can at least account for occlusion. An early study indicated that the region would adapt to a shape regardless of whether it was occluded by vertical bars or not (Figure 1.11A) (Kourtzi and Kanwisher, 2001). However, it would not adapt when the same contours were present in an image but the perceived shape differed (generated by stereoscopically altering the figure-ground relationship of two contiguous shapes; Figure 1.11B). This suggests that LOC may be performing object completion to some extent, and it also supports Kourtzi et al. (2003); again indicating that depth information is only relevant for LOC when it helps distinguish between shapes.



**Figure 1.11.** Shape versus contours. A given object can have the same shape, but different local contours (A), or a different shape with the same local contours (B). The images in B used stereoscopic depth to manipulate the figure (F)-ground relationship. Figure adapted from Kourtzi & Kanwisher (2001).

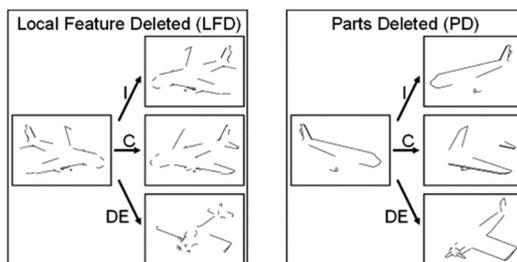
Lerner, Hender & Malach (2002) then explored these completion effects more explicitly, by examining responses to line drawings occluded by vertical black bars (such that 50% of the image was covered, in alternating segments). These images were contrasted with scrambled versions, in which the location of uncovered sections were permuted (creating images that contained the same local details, but that could not support completion effects) (Figure 1.12). Whilst responses in LOC were

higher for whole than occluded objects, the occluded objects nevertheless elicited significantly greater activation when compared to the scrambled versions. Crucially, this was true even for unfamiliar objects, indicating that completion does not rely on prior knowledge of an object's shape. When looking for activation to occluded over scrambled images, the same regions involved in whole object recognition were identified, suggesting LOC may be the source of completion effects.



**Figure 1.12.** Contour completion. If an intact object is occluded by bars, we can still readily perceive its form, demonstrating contour completion. Such an image can be controlled for by permuting the location of non-occluded segments, yielding an image with the same local details but no coherent form. Figure adapted from Lerner, Hendler & Malach (2002).

Further evidence for this comes from a study which used objects with either deleted local features, or deleted parts (Hayworth and Biederman, 2006). Local feature-deleted (LFD) images had regions around alternating vertices removed (e.g. the point where the wings of a plane meet the body). Part-deleted (PD) images had alternating parts removed (e.g. the wings on the plane). The degree of deletion was matched across both image types and complementary versions were also generated, such that superimposing the complementary image on top of the original feature- or part-deleted image would form the whole shape (Figure 1.13).

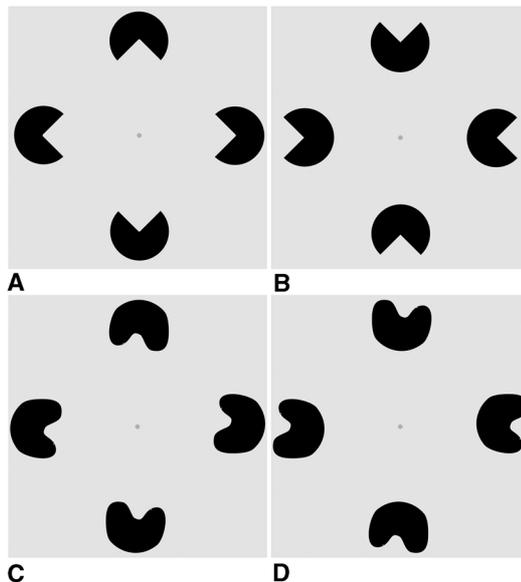


**Figure 1.13.** Deleting local features versus parts. Local feature deleted images have alternating vertices removed, part deleted images have alternating features removed. For both image types, we show an identical but flipped image (I), the complement image (C) and a different exemplar (DE). Figure adapted from Hayworth & Biederman (2006).

These stimuli were then used in an adaptation paradigm, to assess activation in posterior LOC (LO) and pFs. However, these ROIs were defined using a LFD versus scrambled contrast, which potentially biases all activation towards LFD over PD images. This does not necessarily undermine the study, as adaptation looks for relative levels of response, not absolute levels, but a standard (distinct) localiser would arguably have been a better choice. Nevertheless, the results suggested that LO and (to a much greater extent) pFs would adapt to identical but not different LFD and PD images. Critically, adaptation across *complementary* images was only found for LFD, but not PD images (which elicited a complete release from adaptation). This was true even when analysis was constricted to trials when subjects could distinguish between object exemplars. The LFD/complementary result strongly supports the idea that the LOC performs object completion, as it suggests that at least some neurons responded equivalently to the complementary images even though there was no overlap in their local features. Hayworth & Biederman argue that the PD result implies that the LOC actually responds to the parts of an image, however this seems like too strong a conclusion from their results. Seeing (e.g.) the wings of a plane may allow participants to determine what that object is, but the perceptual experience evoked is still just the wings of the plane. As such, it remains possible that LOC would respond differently to a whole plane than to its component parts (i.e. the whole may be greater than the sum of its parts). Still, the idea that LOC is ‘part sensitive’ was also a potential explanation as for why it responds differently to 3D shaded images than to silhouetted versions of those images, so this hypothesis remains a possibility.

It has been hypothesised that LOC’s role in object completion is actually linked to the identification of ‘salient regions’ (Stanley and Rubin, 2003). The theory is that LOC could be segmenting the visual scene into discrete surfaces (that are not necessarily completely bounded), which are then passed down to lower visual areas for more precise processing (i.e. a top-down effect). To test this possibility Stanley & Rubin used Kanizsa-style square stimuli, which give the appearance of circular discs being occluded by an invisible square; the effect is strong enough to create illusory contours along that square’s edges (Figure 1.14A). A second version of this stimulus was created in which the pacman-shaped discs were rounded off and misaligned slightly, creating a salient region without illusory contours (Figure 1.14B). It was

demonstrated psychophysically, by asking participants whether a dot was inside the illusory contour or not, that this manipulation did essentially eliminate the illusory contours as intended. Control stimuli were created by rotating the pacman discs until they were facing outwards, so the same local details were present without the percept of an enclosed shape (Figure 1.14B; Figure 1.14D).

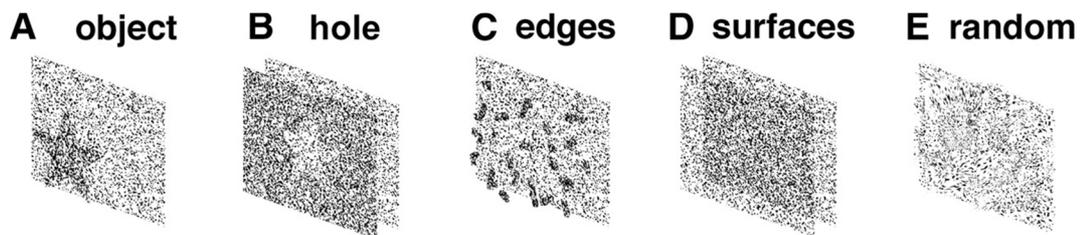


**Figure 1.14.** Kanizsa-style stimuli. (A) An illusory diamond-shape is created by the pacman-shaped discs, however the contours are greatly diminished when the pacman discs are rounded and slightly misaligned (C). Control versions can be created by rotating the pacman discs through  $180^\circ$  (B, D). Figure adapted from Stanley & Rubin (2003).

Illusory contours had previously been found to elicit activity in the lateral occipital region (Mendola et al., 1999), so the standard Kanizsa stimuli were expected to activate LOC; the question was whether stimuli without illusory contours (and just a salient region) would do the same. It was found that they would; both standard and rounded Kanizsa stimuli elicited similar levels of activation in LOC, and both elicited activation that was significantly above that elicited by the control stimuli. As such, it is not simply that LOC was responding to the pacman-shaped discs; the presence of a square region in some form was necessary.

This could support at least two conclusions. One is as Stanley & Rubin hypothesised; that LOC plays a role in identifying salient regions in an image. Alternatively, it could be argued that this is simply evidence that LOC's object completion is relatively powerful, and does not necessarily need clearly defined contours to recognise a shape. We have already seen that inner contours or shading can 'override' outer contours (or at least put them into context). The rounded Kanizsa stimulus still gives the percept of a square, it just has vague boundaries, and so it may simply be this that LOC is responding to.

Another hypothesis to explore is that LOC may be responding to global surfaces (an idea not too dissimilar from ‘salient regions’), and that shape information may simply be useful for their detection. This was tested through the use of planes with different depths (Vinberg and Grill-Spector, 2008). The two planes (front and back, or figure-ground) were defined either stereoscopically, or via motion parallax (in two discrete experiments), and these could be broken down into five subcategories that differed in the form of the front plane (in all cases the back plane was just a flat surface). The first two categories formed shapes; one with the area of the front plane outside the shape cut away, the other with the area inside the shape cut away (such that both shapes had the same external contours). This gave the perceptual experience of either a solid object or a hole in a flat surface respectively (Figure 1.15A-B). The remaining three categories were disconnected edges (created by scrambling the contours of the shapes), a global surface (essentially just a flat plane) and random noise with no coherent structure (control stimulus) (Figure 1.15C-E).



**Figure 1.15.** Objects versus holes. By manipulating coherent motion (or stereoscopic depth; *not shown*) and thus figure-ground, you can define an object (A), a similarly shaped hole (B), edges (C), a sheet surface (D) or random noise (E). Figure adapted from Vinberg & Grill-Spector (2008).

In both experiments, LOC responded significantly more to both objects and holes over random noise. Critically, the difference between object and hole stimuli was also significant, with LOC preferring objects. This indicates that LOC may be sensitive to border ownership, and that outer contours alone are insufficient to elicit maximal activation. LOC showed no significant preferences to edges or surfaces over noise. The former result is unsurprising, given that line drawings > scrambled lines have been used to localise LOC in the past (e.g. Hayworth and Biederman, 2006), however the latter result helps rule out the theory the LOC is generally sensitive to global surfaces. It appears as though some structure (in the form of an object or shape) is necessary to elicit activity in this region. Note that these results

are not simply due to lower-level visual areas, as V1-V3 showed no discrimination between stimuli.

The common question throughout this section has been whether LOC's role is to detect and process outer contours alone. The answer to this question appears to be no. We have seen that inner details (such as shading, or contours that provide depth cues) can cause LOC to both distinguish between two shapes when their outer contours are the same, and treat two shapes as identical when the outer contours are different. Furthermore, LOC appears to identify shapes even when parts of that shape's boundary are occluded, deleted or illusory, indicating that a complete outer boundary is unnecessary for successful object processing. Still, some form of boundary is necessary, as the region will not respond to global surfaces or the control/rotated Kanizsa stimuli from Stanley & Rubin (2003). Furthermore, the section on cue invariance demonstrated boundaries alone could elicit LOC activity. This raises the possibility that LOC may play a dual role. One task may be some form of figure/ground segmentation, in which outer boundaries should suffice. The other may be more subtle object discrimination, in which case details like 3D shading will become necessary. It is unknown whether these two tasks could be performed simultaneously by the same population of neurons, or whether different neurons are recruited for each with the information converging at some later point. Either way, this is still in line with LOC as an intermediate stage in the object processing system; helping to segment and identify objects from the visual scene.

#### **1.4.4 Perceptual Experience of Shape**

One detail present throughout the studies examined so far is that the objects or shapes were detectable by participants. This may seem obvious or perhaps trivial, however neural activity does not necessarily depend upon conscious perception (at least in lower-level visual areas). For example, activity in V1 can predict the orientation of a stimulus, even when that stimulus is rendered invisible to the participant through masking (Haynes and Rees, 2005). This section will address the question of whether detectability is a necessary condition for activity in LOC, and if so, how closely does LOC's activity correspond to our everyday perceptual experience of shape?

#### 1.4.4.1 Detectability

A number of early studies demonstrated that activity in LOC is at least linked to recognisability. First, there is the original LOC study by Malach et al. (1995), which showed that blurring a pixelated image (increasing recognisability due to the Lincoln illusion) caused corresponding increases in LOC activity. This idea was then explicitly tested with a masking study (Grill-Spector et al., 2000), which used image presentations of varying durations to manipulate chance of recognition/detection. This produced non-linear (correlated) increases in both recognisability and LOC activity as presentation duration increased. Furthermore, when participants were trained to recognise images after just 40ms exposure, those trained images elicited significantly higher activation in LOC when compared to novel images. V1 showed no such change in response for trained versus novel images, suggesting that the results were unlikely to be due to greater attention towards trained images (as attention might be expected to modulate V1 responses; Kanwisher and Wojciulik, 2000).

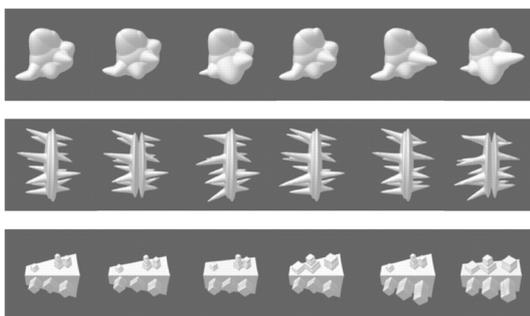
Similar results were found for a more fine-grained definition of recognition by Bar et al. (2001). They had participants rate their degree of recognition via a button press (1–4), with ‘4’ representing successful recognition. A contrast between stimuli with a ‘4’ response versus a ‘3’ response identified an area in ventral-temporal cortex (co-ordinates within the area typically defined as LOC) and an area near the fusiform gyrus. Whilst the lack of localiser makes it difficult to draw conclusions about LOC specifically, the results are in line with LOC activation being tied to the recognisability of objects.

More compelling evidence for the importance of detectability to LOC came from a study based upon the evidence from visual form agnosia and patient DF. DF has bilateral damage in regions corresponding to LOC in healthy participants. Critically, this impairs her ability to make perceptual judgements about objects (e.g. recognising line drawings), but it does not impair her ability to perform motor actions towards objects (James et al., 2003). This in itself (alongside TMS research in healthy participants; Pitcher et al., 2009) is good evidence for a *causal* role of LOC in object recognition, but it also supports the two stream hypothesis (Goodale and Milner, 1992). This is the proposal that the dorsal visual stream is involved in

visually guided actions, and the ventral visual stream (including LOC) is involved in object perception and recognition.

Fang & He (2005) used this idea, alongside binocular rivalry, to show that conscious detection of an object is necessary for activation in the ventral stream (LOC, temporal areas), but that this was not necessarily true for activity in the dorsal stream (V3A/V7/intraparietal sulcus). Two images (objects and textures) were presented simultaneously (via red-green anaglyph eyeglasses), with the object having low luminance and contrast compared to the image in the dominant eye. This ensured that it would be suppressed (invisible) or only just visible when a texture or blank field was presented to the dominant eye respectively. Whilst dorsal areas showed no significant effect of condition (invisible, visible), the ventral regions showed strong activity to visible objects, which was reduced almost to baseline when those same objects were rendered invisible. A second experiment replicated this effect, showing that ventral areas only responded to tools and faces when visible, but dorsal areas would respond significantly more to tools over faces (in line with visually-guided action hypothesis) whether visible or not. As the object was actually present throughout the study (just suppressed), this is a strong indication that conscious detection is indeed necessary for LOC activation.

Finally, Op de Beeck, Baker, DiCarlo, & Kanwisher (2006) found an intriguing effect with respect to training. Grill-Spector et al. (2000) already demonstrated that recognition training could increase LOC activation to previously undetectable objects. Op De Beeck et al. extended this by (extensively) training participants to recognise just one out of three classes of novel stimuli (Figure 1.16).



**Figure 1.16.** Novel object stimuli. Three classes of object stimuli ('Smoothies', 'Spikies' and 'Cubies') used for discrimination training (see main text). Figure adapted from Op de Beeck et al. (2006).

Scans before and after training showed that this caused significant increases in LO (particularly right hemisphere) and right pFs activity for the trained object class. This initially seems to clash with the idea that LOC is not sensitive to object familiarity,

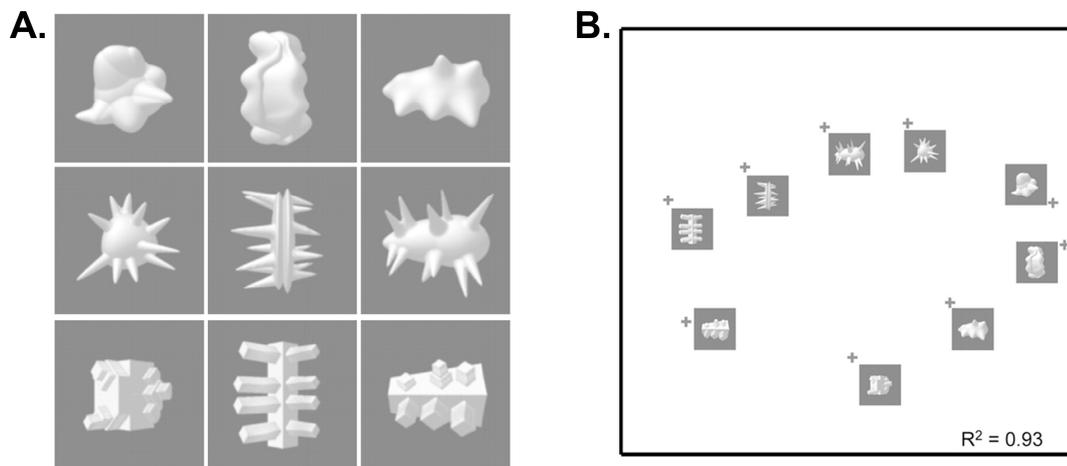
as training should increase familiarity with a set of objects and here we see that this leads to a corresponding increase in LOC activity. However, the differences within a given object class were relatively subtle (hence necessitating training), so perhaps repeated presentation could cause adaptation. It is therefore possible that through training, participants became better able to discriminate within an object class, and this was sufficient to release LOC from adaptation when that class of objects were presented post-training (however, it is unclear why this effect should be lateralised to the right hemisphere). If this explanation holds true, then it suggests that activity in LOC is not only tied to the detection of objects, but also discrimination between objects. Given that training improved the participant's ability to perceptually discriminate these objects, it seems plausible that activity in LOC might actually correspond to our perceptual experience of shape.

#### 1.4.4.2 Perceptual Experiences

There is evidence that LOC's activation is linked to correct discrimination. Williams, Dang & Kanwisher (2007) utilised three sets of novel stimuli, which participants had to categorise. Pattern classification showed correlations between the activity in V1 for a given category across odd and even runs, regardless of whether participants were successful at categorisation or not. Critically, the same was only true for LOC when participants *could* categorise (and therefore discriminate between) the stimuli. This implies that even when lower-level visual areas have detected physical differences between stimuli, this does not necessarily lead the participant to recognise those differences. Instead, it appears to be activity in LOC that cues this discrimination.

Further support for this idea comes from a study that again used novel stimuli (Figure 1.17A), this time with nine discrete object classes (Op de Beeck et al., 2008b). First, a behavioural task had participants rate the similarity between all pairs of stimuli. The distributed pattern of activity in LOC corresponding to each shape was then assessed, whilst participants performed a task that maintained attention. Multidimensional scaling suggested that similarity judgements were primarily being made on the individual features (e.g. are protrusions curvy or spikey; Figure 1.17B), whereas physical (pixel-wise) similarity was mostly based upon the overall shape profile (e.g. is the shape elongated or square). It is possible that this is a consequence of using novel shapes, as unfamiliarity may push participants to look for local

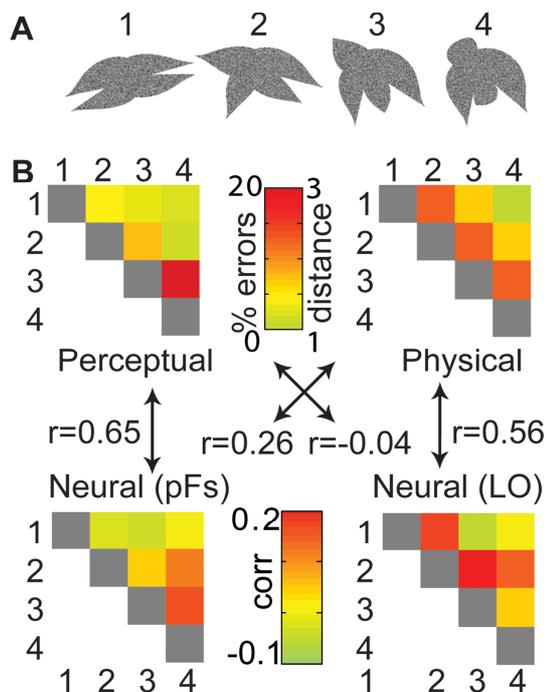
features that aid discrimination. Nevertheless, this result means physical and perceptual similarity could be addressed separately. It was found that LOC's pattern of responses were not linked to physical similarity at all, but were correlated with perceived similarity. Furthermore, both sub-regions of LOC (LO, pFs) showed the same pattern of response, but lower-level visual areas (V1, V2, V4) were not related to perceived shape. Note that these findings were produced in two discrete experiments, each with slightly different tasks, indicating the results are likely to be robust. As such, this lends strong support to the link between LOC and subjective shape perception.



**Figure 1.17.** Neural similarity in LOC. (A) Nine example stimuli, varying in shape envelope (columns) and shape features (rows). (B) Multidimensional scaling (MDS) solution based upon LOC neural data; shapes that cluster closer together (and are therefore more similar) tend to share similar shape features. Figure adapted from Op de Beeck, Torfs & Wagemans (2008b).

However, whilst Op de Beeck et al. (2008b) found similar patterns of response in LO and pFs, Haushofer, Livingstone & Kanwisher (2008b) actually identified a dissociation between these two regions. The general approach was similar between the two papers; using a novel stimulus set, compare the distributed pattern of neural responses with perceptual and physical similarity measures, to see which is playing the greater role. Critically, each paper used different measures of similarity. First, for perceptual similarity Haushofer et al. used discrimination performance, with the rationale that perceptually similar images will be confused as identical more often than perceptual dissimilar images. This method is arguably less susceptible to bias or confounding influences than asking participants to explicitly rate the similarity of two images. Secondly, for physical similarity Haushofer et al. measured how

separated their stimuli (Figure 1.18A) were along (parametrically varied) dimensions of aspect ratio and skew (two properties that are likely to be of relevance to the visual system; Kayaert et al., 2005). Again, using a physical similarity measure based upon perceptually relevant details is arguably better than the pixel-wise similarity approach from Op de Beeck et al, which could be unduly influenced by accidental properties of the image. It is notable that Haushofer et al. also found their physical measure correlated slightly with perceptual similarity, which was not found in Op de Beeck et al; this corroborates the idea that the physical similarity measure chosen is perceptually relevant. Using these stimuli, it was found that LO was correlated with physical similarity ( $r = .41$ ), but not perceptual similarity ( $r < .01$ ). In contrast, pFs correlated with perceptual similarity ( $r = .35$ ) but not physical similarity ( $r = .10$ ) (Figure 1.18B). Furthermore, the profile of activation was relatively consistent between participants in LO, but highly variable in pFs. This fits with their findings as physical similarity is based upon objective measures, whereas perceptual similarity is likely a more subjective judgement. In sum, these results still support the idea of LOC as a region tied to perceptual awareness, but they also indicate that the more subjective, perceptual experience of objects might arise in separate regions to those that just code general object shape. The results also illustrate the importance of careful choice in measures when assessing similarity.



**Figure 1.18.** Perceptual versus physical shape similarity. (A) Stimulus set varying in parameters of aspect ratio and skew. (B) Perceptual, physical and example neural (pFs + LO) similarity matrices. Note that LO appears to reflect physical similarity (distance in aspect ratio/skew space) whereas pFs appears to reflect perceptual similarity judgements. Figure adapted from Haushofer, Livingstone & Kanwisher (2008b).

Finally, the studies above have predominantly used novel, abstract shapes, instead of more familiar objects. This is often placed as an advantage (e.g. Haushofer et al., 2008b), as it avoids confounds such as semantics when judging similarity. However, it could be argued that making similarity judgements on these fairly arbitrary shapes forces unusual processing strategies, potentially rendering the results irrelevant to real world object processing. This does not appear to be the case. Using photographs of real animals in an MVPA study (similar methods to above), Weber et al. (2009) found that activity in bilateral ROIs likely corresponding to LOC were significantly correlated with similarity judgements. It should be noted that they originally asked participants for *conceptual* similarity judgements, however they later found that these were highly correlated with visual similarity judgements ( $r = .94$ ), perhaps because images override more subtle semantic associations. As in Op de Beeck et al., no ROIs correlated with pixel-wise physical similarity ratings, again arguably demonstrating that this is an inadequate measure of physical similarity.

In summary, there is strong, converging evidence that LOC is the region where our perceptual experience of objects originates. If we do not perceive an object then it does not elicit LOC activity, even when physically present. If we fail to discriminate between objects, the activity in LOC cannot be used to discriminate between those objects. Finally, if two objects are judged to be similar, then this is reflected in similar activation distributions with LOC. It remains to be seen whether the LO/pFs dissociation between physical and perceptual similarity holds true. Nevertheless, this finding could explain the studies showing increasing invariance as you move anteriorly from LO to pFs. We perceive objects to be the same whether they are in different locations or different sizes, so it would make sense that a region linked to perceptual similarity would show some invariance to those manipulations.

## 1.5 Concluding Remarks

Malach et al. (1995) initially proposed that the Lateral Occipital Complex (LOC), a region which preferentially responds to objects over scrambled objects, could be a good candidate for an intermediate object processing centre. Throughout this chapter we have discussed and critiqued evidence that (generally) appears to support this assertion. It is a region that can be reliably localised both between and within participants, and it appears to respond to object shape regardless of the cue that is used to define it. It also seems to show slight invariance for manipulations in size

and viewpoint. Furthermore, LOC does not simply respond to the outer contours of a shape, but can make use of 3D depth cues like shading to help differentiate between objects. It is also able to respond to shapes when occluded, which corresponds to our ability to perceive the full shape of an object, even when it may not be in full view. Finally, there is substantial evidence supporting the idea that LOC may be the region where our perceptual experiences of object recognition first emerge.

There are still multiple lines of enquiry worth pursuing, and these form the foundation for this thesis. In particular, the relationship between retinotopic areas LO-1 and LO-2 with LOC is currently unclear. Whilst we have evidence for overlap between these regions, this does not necessarily imply shared functional specialisation. Indeed, there is markedly little evidence on the functional properties of LO-1 and LO-2 more generally, and so this topic is ripe for exploration. There is also the wider question regarding the nature of the object representation within the LOC itself. In the empirical chapters to follow, we will ask both what that representation is, and how does it arise from processing earlier in the visual stream?

## Chapter 2. Methods Review

### 2.1 Overview

The aim of this chapter is to provide an overview of the methods that will be used throughout this thesis, as well as the underlying rationale behind those methods. Most of the empirical chapters to follow share a common framework. For example, the focus on visual areas LO-1 and LO-2 (Larsson and Heeger, 2006) means that all chapters will employ retinotopy to identify those regions. As such, describing the methodology here should avoid some redundancy. The chapter will be broken down into seven broad sections:

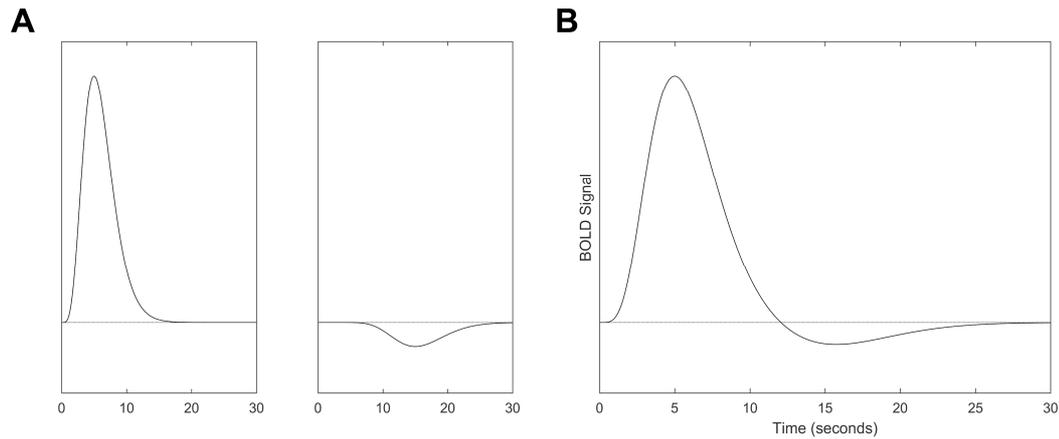
- 2.2. **Functional Magnetic Resonance Imaging (fMRI):** this section will provide a brief overview of fMRI, most notably the hemodynamic response function which governs a large number of the design decisions in our experimental paradigms.
- 2.3. **Structural Scans:** this section will describe the process of acquiring high-resolution structural scans, and segmenting them into grey/white matter so data restricted to activity derived from cell bodies can be analysed.
- 2.4. **Experimental Designs:** this section will discuss the merits of two different fMRI experimental paradigms; block and rapid event-related designs. It will also describe how data generated from such designs will be analysed.
- 2.5. **LOC Functional Localiser:** this section will describe the stimuli used to define the Lateral Occipital Complex, how those stimuli are presented and analysed, and the various ways regions of interest (ROIs) can be derived from those data.
- 2.6. **Retinotopic Mapping:** this section will describe the stimuli used for retinotopic mapping, how those stimuli are presented and analysed (travelling wave analysis), and finally how ROIs are derived from those data.
- 2.7. **Radial Frequency (RF) patterns:** this section will introduce RF patterns; a set of shape stimuli that were introduced in the psychophysics literature (Wilkinson et al., 1998), and have been used extensively throughout this thesis.
- 2.8. **Fourier Descriptors (FDs):** this section will introduce FDs (Zahn and Roskies, 1972), which allow you to describe and manipulate shapes in the frequency domain, providing some useful techniques for stimulus control.

## 2.2 Functional Magnetic Resonance Imaging (fMRI)

### 2.2.1 The BOLD response

Arguably the most fundamental concept for fMRI is that it cannot measure neural activity directly. For neurons to fire they require energy, and this energy must be replenished in the form of increased blood flow (due to the increased metabolic demand for oxygen). As such, a region of active neurons will have greater general blood flow than regions of inactive neurons. As blood flow increases there will be a rise in the ratio of oxygenated to deoxygenated blood, before levels gradually return to the initial state. Critically, fMRI can discriminate between oxygenated and deoxygenated blood, therefore this blood flow can be detected (Logothetis and Wandell, 2004). This change in blood flow is termed the hemodynamic response, and it is detected with a Blood-Oxygen-Level Dependent (BOLD) contrast (Ogawa et al., 1990). In sum, measuring the BOLD response is a means to measure neural activity by proxy.

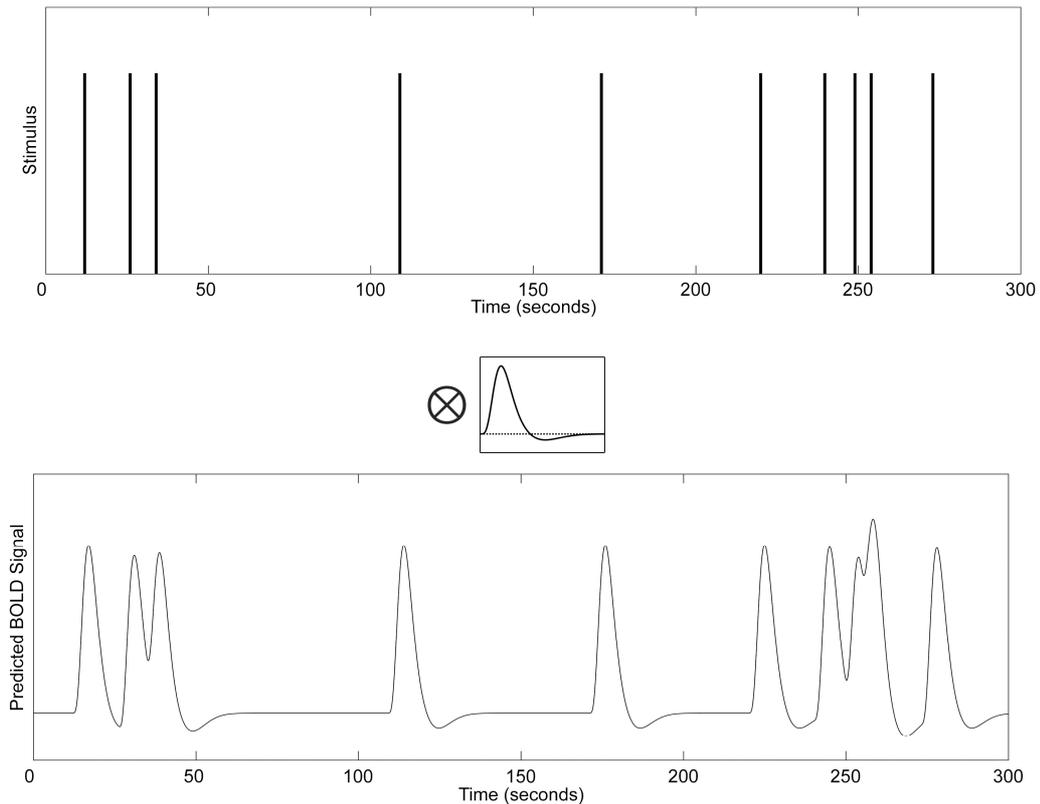
As we are not measuring neural activity directly, there is an inherent lag (temporal delay) in fMRI due to the nature of the hemodynamic response. The BOLD response is not instantaneous; instead it follows a characteristic pattern over a period of several seconds. It first starts to rise in response to neural activity after 1-2sec, before peaking after around 4-6sec. It then begins to fall, dipping slightly below baseline (undershoot), finally then returning to baseline after around 16-18sec. However, the full process from start to complete return to baseline can take up to 25-30 seconds. This pattern can typically be well approximated by a double gamma hemodynamic response function (HRF), shown in figure 2.1.



**Figure 2.1.** Double gamma hemodynamic response function (HRF). Through the combination of two gamma functions (A), one to model an initial peak then another to model the subsequent undershoot, we can create a good estimate of the HRF (B).

To successfully model the hemodynamic response across an experiment we assume that the BOLD response is a linear system. A critical distinction is needed here, as there are two discrete processes at work and only one of which is assumed to be linear. First, when stimuli are presented this evokes neural activity, and this activity is highly *non-linear*. That is, doubling the intensity of the stimulus will not (necessarily) double the resultant neural activity. However, we assume the BOLD response to that neural activity *is* linear; such that a twofold increase in neural activity would elicit a twofold increase in the BOLD response.

In a linear system, if you know the response generated from a unit impulse (a theoretical input comprising perfect stimulation with infinitely short duration) then you can model the expected response to any series of events (by modelling those events as a series of unit impulses). The double gamma HRF is exactly that; the expected response to a unit impulse. As such, through the assumption of linearity (an assumption that largely holds true for present purposes; Hansen et al., 2004) we can model any experimental design and convolve it with the double gamma HRF, producing a predicted BOLD response (Figure 2.2). By comparing this predicted BOLD response to the measured BOLD response acquired during an fMRI scan we can start to tease apart the neural activity elicited by our chosen stimuli.



**Figure 2.2.** Predicting neural responses. The top panel shows a set of ten events at randomly spaced intervals across 300sec. By convolving this with the double gamma HRF we can predict the expected BOLD response to this set of stimuli (bottom panel), which can be compared to actual neural responses. Note that the height of the predicted signal depends heavily on the inter-stimulus interval between events; an important consideration when calculating percent signal change.

### 2.2.2 Measuring the BOLD response

So far we have discussed how a BOLD response is generated after stimulus presentation, however this response also has to be *measured* so that we can compare it to our predictions.

Fundamentally, responses in MRI and fMRI are generated by exploiting magnetic fields. Consider a three-dimensional axis with the  $z$  (longitudinal) dimension aligned along the bore of the scanner. The  $x$  and  $y$  dimensions would therefore be the transverse plane. When a participant enters the strong, static magnetic field of a scanner, hydrogen nuclei (protons) essentially act as bar magnets and tend to align themselves with that field (i.e. in the longitudinal direction), reaching equilibrium.

Critically, the protons can either be aligned parallel to that field in a relatively low energy stable state, or they can align antiparallel to the field in a higher energy less stable state. This stability discrepancy results in a slight bias towards the parallel state, resulting in net magnetisation in the longitudinal direction.

The picture is complicated to some extent, as protons have both mass and spin, giving them angular momentum. This means that they do not simply align with the longitudinal axis/magnetic field quite as a bar magnet would, instead they precess about it (i.e. they rotate in a gyroscopic fashion about the direction of alignment). In theory, this could cause magnetisation in the transverse plane, however the large number of protons/spins means that this tends to cancel out (specifically, there is no spin coherence).

Now, if brief radio frequency (RF) at the resonant frequency of the nuclei is applied, this disturbs the equilibrium by providing the nuclei with energy. As a result, those protons in the low energy state gradually switch to the high energy state, reducing net magnetisation along the longitudinal axis. By applying just the right amount of energy (specifically, a 90-degree excitation pulse), a balance can be reached where there are the same number of protons in the low and high energy states. This cancels out net magnetisation along the longitudinal axis. Crucially, this also brings the spins of those protons which have switched states into phase, resulting in spin coherence and therefore (measurable) net magnetisation in the transverse plane.

The protons are now in an overall less stable state, and so over time the supplied energy is released as equilibrium returns. This process is termed relaxation, and there are two discrete types;  $T_1$  and  $T_2$ . First, due to the inherent instability of the high energy (antiparallel) state, protons will naturally tend to revert to the lower energy state over time. This process is known as  $T_1$  recovery, as it results in a gradual *restoration* of the net magnetisation in the longitudinal axis. When protons flip from the high-to-low energy states, they also release RF energy at the same frequency as the previously applied pulse; therefore, the same coil used to apply this energy can also be used to detect it upon release. The second type of relaxation occurs in the transverse plane. Whilst the spin states are initially brought into coherence by the excitation pulse, over time interactions or interference between spin states causes the phases to gradually lose coherence. This process is termed  $T_2$  decay as it results in a

*loss* of net magnetisation in the transverse plane. Note that local inhomogeneities in the magnetic field can also cause phases to lose coherence, this effect combined with the effects of spin interactions ( $T_2$  decay) is referred to as  $T_2^*$  decay, and will prove vitally important shortly.

Critically, precession frequency is proportional to magnetic field strength. Therefore, by applying magnetic gradients on top of the static field you can systematically vary the resonant frequencies of the hydrogen nuclei. This means that discrete populations of nuclei will be excited by the RF at any given time, dependent upon the applied magnetic field gradients. By measuring responses (released energy), information can be collected from just that discrete population, allowing spatial specificity.

Finally, as discussed previously fMRI exploits differences between oxygenated and deoxygenated blood to detect the BOLD response. To elaborate, blood contains iron, and iron has slightly different magnetic properties dependent upon whether it is in the presence of oxygen or not. Specifically, deoxygenated blood has greater magnetic susceptibility than oxygenated blood. This means that deoxygenated blood will interfere with the magnetic field, at least relative to oxygenated blood, causing local inhomogeneities in that field. From earlier, we know that such inhomogeneities can produce some degree of transverse relaxation, known as  $T_2^*$  decay. Therefore, by exploiting magnetic gradients (for spatial specificity) alongside a  $T_2^*$ -weighted image, we can measure the BOLD response across the brain.

### **2.2.3 Scan Parameters**

To conduct an fMRI scan, decisions have to be made regarding the scan parameters to use. Scan parameters are essentially a set of instructions determining how that BOLD response is measured, and they can have a critical impact in the quality of the collected data.

The first parameters to examine regarding this process are repetition time (TR) and echo time (TE). Repetition time is the duration between successive RF pulses. Intuitively it may seem as though a fast TR is desirable, as a faster sampling rate should increase temporal resolution. However, if the TR is too fast there will not be sufficient time for recovery from successive excitations, reducing the signal-to-noise ratio (SNR). Longer TRs also allow for more slices to be collected, increasing brain coverage (but also resulting in longer scans). The inherently slow nature of the

hemodynamic response must also be considered, as it means an increased sampling rate does not necessarily imply increased temporal resolution. As such a balance is needed, typically a TR of 2-3sec was used depending upon the design paradigm employed (see *Experimental Designs*, section 2.4). Note that the excitation recovery consideration has a secondary implication. At the start of the scan there will be maximum possible alignment, this alignment will be disrupted through RF excitation before returning to maximum alignment after some period of time. However, with a short TR the alignment will be disrupted through excitation before maximum possible alignment could be reached. This means that signal will be strongest at the very start of the scan, dropping off slightly until equilibrium is reached. As such, we typically acquire (and discard) dummy volumes at the start of each scan to account for this process. The second parameter, TE, is the time between the RF pulse and response measurement. This was typically set as 30ms across all our functional experiments. A related parameter is flip angle; conceptually this is the angle to which excited nuclei are actually flipped as a result of an excitatory RF pulse. This was typically set to the default of 90°, however in the last two experiments (Chapter 7) it was set to the Ernst angle. This is the (TR-dependent) optimal flip angle and so should help optimise SNR.

The final set of parameters we will discuss relate to spatial resolution; field of view (FOV), matrix size and voxel size. These parameters are all interrelated such that changing one affects all others. First, FOV is the size of the image to be acquired and that is split into a grid based on the matrix size. For example, you may have a FOV of 192mm and matrix size of 96x96x39. This implies 39 slices with an inplane voxel size of  $2 \times 2 \text{mm}^2$  ( $192/96 = 2$ ). Last there is slice thickness, which intuitively determines the thickness of said slices. We typically set this to match the in-plane voxel size, increasing it on occasion to account for shorter TRs (which would normally reduce the brain coverage as a result of fewer slices being possible to acquire in a shorter period of time). Note that it is also possible to introduce a slice gap, which increases spatial coverage at the cost of losing information between slices, however we did not need to utilise slice gaps for this thesis. Again, determining the resolution requires a balance. A smaller voxel size is preferable as this increases spatial resolution; particularly desirable when you aim to explore smaller ROIs such as LO-1 or LO-2. However, smaller voxels imply fewer sampled

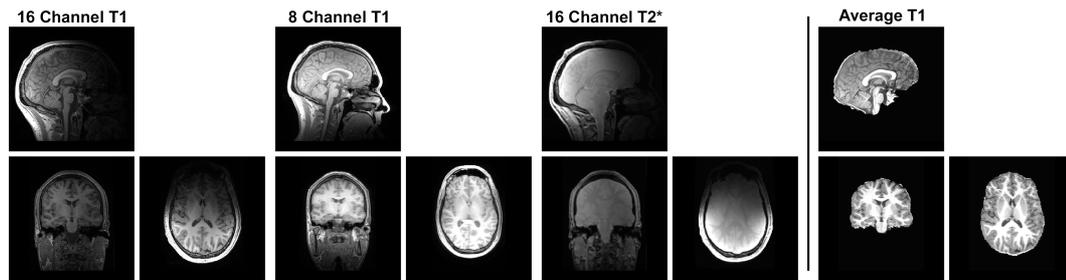
neurons per voxels, therefore less blood flow (and fewer protons) to measure, which reduces SNR. Voxel sizes varying from  $2 \times 2 \times 2 \text{mm}^3$  to  $3 \times 3 \times 3 \text{mm}^3$  were used in the functional paradigms described later.

## 2.3 Structural Scans

### 2.3.1 Data acquisition

One of the first steps performed on new participants is the acquisition of high-resolution structural MRI scans. As discussed above, for fMRI we typically look for voxel sizes in the region of  $2\text{-}3 \text{mm}^3$ , a relatively coarse resolution. Furthermore, over the course of a scan multiple volumes are collected (one every TR), and those volumes do not necessarily encompass the whole brain depending upon the scan parameters collected. All this means that fMRI can provide useful insights into *functional* properties of the brain, but limited information regarding *structure*. The issue with this is that it means the anatomical specificity of any detected activation can be hard to pin down. To address this, we acquire separate structural scans, which typically collect just one volume over the course of a scan with a much smaller voxel size (typically  $1 \text{mm}^3$ ). Critically, because both the functional and structural scans share the same underlying anatomy, those functional scans can be mapped (aligned) onto the higher resolution structural images, making it easier to identify where activation is situated. Notwithstanding, the different acquisition techniques used in structural MRI and fMRI can result in distortions between the spatial frames (e.g. at the anterior temporal lobe), which sometimes need to be accounted for.

For our purposes, we typically collected a comprehensive set of structural scans on each participant. Specifically, we collected **three 16-channel T1-weighted anatomical images** (TR = 7.8ms, TE = 3.0ms, TI = 600ms, voxel size =  $1 \times 1 \times 1 \text{mm}^3$ , flip angle =  $12^\circ$ , matrix size  $256 \times 256 \times 176$ , FOV = 256mm), **one 8-channel T1-weighted anatomical image** (TR = 7.8ms, TE = 2.9ms, TI = 450ms, voxel size =  $1.13 \times 1.13 \times 1 \text{mm}^3$ , flip angle =  $20^\circ$ , matrix size  $256 \times 256 \times 176$ , FOV = 290mm) and **one 16-channel T2\*-weighted fast gradient recalled echo scan** (TR = 400ms, TE = 4.3ms, voxel size =  $0.375 \times 0.375 \times 2 \text{mm}^3$ , flip angle =  $25^\circ$ , matrix size  $512 \times 512$ , FOV = 192mm). Examples of these scans can be seen in Figure 2.3.



**Figure 2.3.** Example Anatomical Images. The two sets of scans on the left show T1-weighted anatomical images generated using either a 16- or 8-channel head coil. Note that data collected using the 16-channel coil shows a substantial gradient, which can be partially corrected for using the 16-channel T2\* scan. Once ‘combined’ (see main text), we produce an Average T1 for further processing.

We will briefly explore the differences between these scan types. First, at the York Neuroimaging Centre we can acquire data using two different head coils, an 8-channel phase-array head coil and a 16-channel half-head phase-array head coil. The 16-channel coil has numerous advantages over the 8-channel coil and so was used for all functional MRI throughout this thesis, as well as the bulk of the structural scan sessions. Specifically, it has twice the number of channels, and being a half-head coil these are all situated about the posterior part of the brain yielding greater SNR in visual cortex. A tertiary benefit is that by only covering the back of the head, there is greater visibility for the participant – useful for visual experiments. However, the increased resolution towards the back of the brain comes at the cost of poor signal towards the front. This results in quite a strong gradient in signal when moving posterior-to-anterior through the brain. We can correct for this relatively well using the T2\*-weighted scan (which provides a measure of this gradient bias as  $T_2^*$  decay is associated with magnetic field inhomogeneities), nevertheless it can make certain processes difficult (namely brain extraction). To address this, we also acquire one 8-channel T1-weighted scan, which has broadly uniform signal across the whole brain. The uses of this scan are discussed below (see Data Analysis, section 2.3.2). It is worth noting that the 16-channel head coil has a second, less obvious disadvantage. As it only covers the back of the head, participants can only be ‘secured’ using a strap across their forehead (versus the 8-channel coil in which their whole head is secured inside). This means that 16-channel scans can be more susceptible to participant movement (an unwanted source of noise), and care was

taken to ensure participants knew to remain as still as possible throughout their scans.

The second part to mention regarding our structural scans is that we refer to them as T1-weighted. In section 2.2, we briefly discussed the theory through which MRI and fMRI actually detects signal, including the process of relaxation (nuclei returning to alignment after excitation from an RF pulse). Critically, we discussed two types of relaxation,  $T_1$  and  $T_2$ , and different tissue types have different relaxation properties. Therefore, by biasing detection towards one relaxation type or another (through TR/TE choice), we can highlight different tissue types. As an example, at magnetic field strength 1.5T we find the following  $T_1$  and  $T_2$  properties (Huettel et al., 2009):

- White matter:  $T_1$  600ms,  $T_2$  80ms
- Grey matter:  $T_1$  900ms,  $T_2$  100ms
- Cerebrospinal fluid:  $T_1$  >2000ms,  $T_2$  2000ms

Based upon this, by choosing a short TR and short TE (T1-weighted scan) we should get strongest signal in white matter (i.e. white matter will appear ‘brightest’), as this has the shortest relaxation time. This will be followed by grey matter, and then areas with water (e.g. CSF) should be dark. Such a scan is excellent for highlighting anatomy; hence we use T1-weighted structural scans.

### 2.3.2 Data Analysis

As described above, in our structural scanning process we acquire five discrete scans, however we only actually need one high-resolution image of the brain. The process of generating this image will now be described. Note that most processes will use FSL (<http://fsl.fmrib.ox.ac.uk/fsl>), a suite of fMRI analysis tools. First, the reason we acquire three 16-channel T1 scans is that by averaging we can increase SNR (although this does have diminishing returns, for example averaging two scans does not double SNR, it only increases it by  $\sqrt{2}$  at the cost of twice the scan duration). To average, we do the following:

1. First we do a preliminary brain-extraction using FSL’s BET (Smith, 2002) on all three T1 scans. This is also referred to as skull stripping, as it extracts just the brain from the image. By doing this we ensure alignments are conducted on the brain matter itself, which naturally is the region in which we want the

best possible alignments to maximise SNR. Note that as we are using 16-channel images here, the BET will typically clip large parts off the front of the brain, but it serves for the current purposes.

2. Next, we align all three T1 images to the first T1 scan acquired, using FSL's FLIRT (Jenkinson and Smith, 2001; Jenkinson et al., 2002).
3. Finally, we can average the three scans together using *fslmaths*. The result of this process will be referred to as the 'Average T1'.

Next, we want to perform a more accurate brain-extraction, as the inhomogeneity gradient in the 16-channel images makes initial extraction difficult.

4. First, we align the 8-channel T1 to the Average T1.
5. We then skull strip the 8-channel in 'Average T1-space', saving out a 'brain mask'. This is essentially just a binary image detailing where the brain structure is in our image.
6. Using this brain mask, we can now successfully skull-strip the Average T1.

The final process is to correct for the inhomogeneity gradient in the Average T1.

7. First, the T2\*-weighted image (which contains an estimate of the magnetic field bias) is aligned to the Average T1.
8. We then apply the same brain mask as before to skull-strip the T2\*-weighted image.
9. Finally, we simply divide the Average T1 by the skull-stripped T2\*-weighted image, largely correcting for the gradient bias inherent in 16-channel images.
10. As an additional step we also threshold the Average T1 at this stage to sensible values, because the divide performed above can sometimes cause spuriously low or high voxels, which may result in undesirable effects on subsequent alignments.

Note that steps 4-10 are actually repeated twice, except that in step 4 of the second run the Average T1 used is the output Average T1 from the first complete run (before then continuing with the Average T1 from step 3). The reason we do this is because a slightly superior alignment can sometimes be achieved by aligning the 8-channel T1 to a gradient-corrected skull-stripped Average T1, resulting in an overall better brain extraction. An example Average T1 can be seen in Figure 2.3.

### 2.3.3 Brain Segmentation

One of the chief reasons we aim for such a high quality anatomical volume is that we then segment it into grey and white matter. Grey matter is of primary interest as it contains the neuron cell bodies and synapses, whereas white matter is largely just axons. As such, by excluding white matter we can reduce our image to just informative voxels. Furthermore, grey matter segmentation allows us to view data on an inflated surface; useful for drawing regions of interest (ROIs) and exploring clusters of activity.

This segmentation follows a two-step process. First, we run an automated segmentation using Freesurfer's (<http://surfer.nmr.mgh.harvard.edu/>) autorecon script, with an additional flag to skip the skull strip option (as that has already been manually performed, as described above). The output of this process is then manually edited and checked for accuracy using ITKGray (<https://web.stanford.edu/group/vista/cgi-bin/wiki/index.php/ItkGray>), a branch of ITKSnap (Yushkevich et al., 2006). To ensure a good segmentation has been achieved, all three axes are examined (axial, sagittal and coronal), spurious volumes are deleted, cavities are filled (as the grey matter should form one 'sheet') and handles (partially-connected grey matter) are manually corrected for.

### 2.3.4 Axial PDs

One final scan type worth mentioning in this section is the 'Axial PD' (example parameters: TR = 2700ms, TE = 34.84ms, voxel size = 1x1x2mm<sup>3</sup>, flip angle = 90°, matrix size 192x192x39, FOV = 192mm). This is another form of structural image weighted based upon proton density (hence PD). As described previously, one of the main uses of the high-resolution structural volume is to align functional data to it, helping to correct for the relatively poor spatial resolution of fMRI. However, due to the coarse resolution of fMRI images, this alignment process can prove difficult. To address this, we acquire an Axial PD scan at the same time and with the same slice prescription (i.e. same region of brain included) as our functional scans, but with higher resolution. As the Axial PD has the same slice prescription as the functional scans, the functional scan to Axial PD alignment is relatively easy. The Axial PD scan can then be aligned to the main structural image, resulting in a two-stage mapping from the functional scan to the main structural.

To facilitate alignments, two steps are performed on the Axial PD. We first use FSL's FAST (Zhang et al., 2001) to estimate and correct for the bias field (inhomogeneity gradient) inherent in 16-channel images. We then skull-strip the image as described previously. This ensures that the Axial PD is in the same 'format' as the Average T1 created earlier.

We should note that instead of an Axial PD another scan type known as a T1-FLAIR is often collected. A T1-FLAIR would result in a slightly easier alignment as it has the same weighting as our T1-weighted structural images. However, perhaps due to a quirk with the York Neuroimaging Centre's MRI machine a T1-FLAIR scan takes considerably longer on a 16-channel coil, hence we acquire an Axial PD. The alignment is still relatively straightforward; it just necessitates the use of an algorithm that works independently of tissue weighting (namely *mutual information* based alignment).

## 2.4 Experimental Designs

One of the most important decisions to make when designing an experiment is the experimental paradigm to use. Due to the inherent lag of the hemodynamic response, it is not possible to just present stimuli on screen haphazardly and record the resultant activation. Instead, careful attention must be paid to the ordering and spacing between stimuli, to ensure that neural signals unique to each stimulus type can ultimately be extracted. In this thesis, two different paradigms have been used; block and event-related. Both will be discussed, followed by discussion of how the resultant data is analysed as this is largely equivalent across paradigms.

### 2.4.1 Block Designs

The first paradigm used in fMRI was the block design. In its simplest form, you present two discrete conditions in sequentially repeated blocks of equal length (i.e. ABAB), then subtract activity elicited by one condition from the other.

The underlying foundation of this paradigm is 'cognitive subtraction'; the idea that by subtracting activity elicited by condition B from activity elicited by condition A, you isolate the neural activity uniquely associated with A. Whilst this technique has considerable merit, yielding valuable insight into the organisation of visual cortex (see for example 'LOC Functional Localiser', section 2.5), it does come with certain assumptions. Most notably is that of 'pure insertion'. The idea is that in any

experiment there should be a set of ‘baseline processes’ (more colloquially, thoughts/feelings etc.) and these act as a *constant* source of noise. If true, this noise should be equivalent across both A and B, so by subtracting one from the other you eliminate it. However, this assumes that the influences of A and B are completely independent not only from those background processes, but also from each other (i.e. it assumes the presence of A cannot alter responses to B). A good example is the object localiser, which typically contrasts objects against scrambled versions of those objects. We assume activation from the ‘objects > scrambled’ contrast is related in some way to the visual processing of object stimuli. However, objects can be named, they may conjure up memories or feelings, they can have affordances (e.g. a hammer can be used as a tool), plus they might simply be more visually engaging than unrecognisable scrambled images. Technically, the identified activity could be due to any of these discrepancies, and so when designing such experiments care should be taken to match conditions in all attributes but the attribute of interest (where possible).

Throughout this thesis, block designs have been used when we want to explore responses to a *class* of stimuli. That is, a set of stimuli that all share some common attribute. Block designs excel at this as they are statistically powerful; by presenting similar stimuli over a relatively prolonged period of time the hemodynamic response should plateau, allowing a good estimate of the BOLD response’s final height for that condition (although there are some caveats, e.g. adaptation may diminish the BOLD response to some extent). In terms of block length, there are three main constraints; we want the BOLD response to peak during a block (~ 4-6sec), and also for it to return to baseline when out of the block (~ 16-18sec). Discrete blocks also need to be as close together in time as possible. In fMRI there are inherent sources of noise (physiological and physical) that act across a scan (i.e. low frequency drift). As such, even two identical blocks may elicit different responses if separated by considerable time. Based upon these considerations, we typically set block length to about 15 or 16sec, dependent on TR (integer multiples of the TR are preferable).

The next consideration is baseline choice; that is, what activity will be considered ‘default’ against which all other activation will be compared. A conventional option is to insert an actual baseline ‘condition’ (e.g. blank screen with fixation only) either between every block (AnBn; n = null) or counterbalanced between sets of blocks

(ABCn, ABnC etc.). The latter is arguably preferable as otherwise null periods comprise 50% of the available scan time, although the former does ensure there is no overlap in the hemodynamic response across conditions. Note however, that a null period is not necessarily true baseline. There are likely many comorbid/secondary responses to active conditions that are not present during rests (e.g. motor responses, directed attention or simply visual stimulation). Similarly, it is likely that additional responses arise during rest periods when the participant is not actively engaged (e.g. mind wandering). This again concerns the assumption of ‘pure insertion’. The alternative is to use an ‘active’ baseline, in which the baseline condition is expected to elicit some activation (specifically, activation you aim to control for). For example, in a localiser you aim to isolate object-specific activity, and so it is preferable to model out the effects of simply perceiving a visual stimulus by using scrambled stimuli as baseline. Different combinations of these three options were used throughout the thesis, depending on design choice (e.g. as number of conditions increases, using null periods between every condition becomes increasingly infeasible). Finally, in cases where more than two conditions were used, counterbalancing becomes important to avoid order effects. Ideally full counterbalancing (all possible permutations) was used, but in cases where this was infeasible we used a balanced Latin square design, this ensures every condition was equally likely after every other condition.

#### **2.4.2 Rapid Event-related Designs**

The (increasingly popular) alternative to a block design is an event-related paradigm. In this paradigm, events (stimulus presentations) occur in isolation over the course of the run. Throughout this thesis, we used this design when we want to measure neural responses unique to *individual* stimuli.

Originally, it was thought that events could only occur every 20 to 30sec or so, to ensure the hemodynamic response had time to return to baseline. This made event-related designs highly inefficient. However, with *optimised* timings there can be much shorter inter-stimulus intervals (in the regions of 2-4sec; hence ‘rapid’), leading to much greater efficiency. This is because the assumption of linearity allows overlapping hemodynamic responses to be teased apart (deconvolved), providing measures of neural activity specific to each event type. However, great care must be taken with timings, as poor design choices can undermine the experiment.

The first timing issue to consider is that of jitter; it is critically important that the inter-stimulus interval varies across the run. To demonstrate why, imagine an event occurring every 2sec, this is going to cause a small periodic rise and fall in the hemodynamic response across time. Critically, we also sample at a periodic rate (based on TR), and it is possible that our measurements may just happen to coincide with the point between that rise and fall, therefore missing the true peak. This would greatly reduce the SNR (e.g. imagine sampling at the zero-crossings on a sine wave; it would appear flat). Through jitter, we minimise the risks of this occurring because the rise and fall of the hemodynamic response will never occur at set intervals.

Perhaps equally important is the issue of counterbalancing. Due to the rapid nature of the stimulus presentation, the hemodynamic responses related to each event will greatly overlap. Due to this, if the events occurred in a pre-set order (ABCABC etc.) it would be impossible to determine whether responses for B were actually elicited by B, or if instead they were residual activation from A. Jitter can address this to some extent as if the activity did originate in A it should be independent of exactly when B occurs, but counterbalancing is still warranted as it also helps control for participant expectations. More formally, counterbalancing helps avoid the issue of collinearity (correlated regressors); uncorrelated stimulus presentation sequences will allow greater power when fitting parameter estimates, increasing SNR. In all event-related experiments to be discussed, both stimulus jitter and counterbalancing was optimised using Optseq2 (<https://surfer.nmr.mgh.harvard.edu/optseq>). This was typically run for 24-48hrs, generating unique, optimal schedules (stimulus presentation orders) for each run of the scan. We chose to use these in order of most-to-least efficient, as participants are most alert in earlier runs.

### **2.4.3 Additional Design Considerations**

There are a number of additional considerations common to both block and event related paradigms.

First, there is run length, which ended up being set based upon pragmatic terms rather than anything more objective. Typically, we find participants start to fatigue after about 40-60mins scanning, putting an upper limit on total scan length.

Theoretically this time could be spent on a single run, but asking participants to maintain attention for such a long period of time without rest would be unreasonable.

In work earlier in the thesis we chose run lengths of 8-10mins, leading to 4-5 runs per scan. However participant feedback suggested such runs felt overly long. Due to this, we later switched to shorter run lengths of about 5mins, leading to 8 runs per scan. In addition, we started letting participants know how many runs they had left after each run. Feedback on this switch was positive, suggesting the scan felt much shorter despite being the same overall length. This demonstrates how relatively minor design changes can influence the comfort of the participant, and hopefully the quality of the resultant data. Shorter runs also have a secondary advantage. On occasion runs need to be removed due to movement or other outside influences, having a greater number of shorter runs means less data is lost if this happens.

The final consideration is the task we ask participants to perform. In a given scan we want two things from the participant; stable fixation and attention to the stimuli presented. At the very least the task should not interfere with these conditions, at best it should facilitate them. We can rule out the red dot task, in which the participant responds when the dot falls within the stimulus, as this may encourage eye movements towards it. We can also rule out tasks at fixation (e.g. respond when fixation cross changes colour), as this directs attention away from the stimuli. Perhaps the ideal task is the one-back task, in which participants respond when a given stimulus is presented twice in succession. This encourages participants to actively attend to and remember the stimulus, without necessitating attention away from fixation (as long as stimuli are presented centrally). As such, this was the most common task used throughout the thesis. However, in rapid event-related designs we found the one-back task may not be ideal; the sporadic nature of events means sequential repeats are both rare and may occur far apart in time, so they can be easily missed. As such, we also tried an odd-ball task, in which participants respond when they see some predefined atypical stimulus type. This does introduce some confounds, as there will be additional activation elicited by the detection of and response to the oddball. However, oddball timings can be manipulated independently from stimulus timings, so they can be modelled out to some extent. As such, this task may be preferable in cases where the one-back task is not appropriate.

#### **2.4.4 Data Analysis I**

All fMRI data is analysed using FEAT (Worsley, 2001), part of the FSL (<http://fsl.fmrib.ox.ac.uk/fsl>) suite of fMRI analysis tools. This was first performed at

the level of the individual run using a common procedure across both paradigms. Due to the complex nature of fMRI, a large number of steps (smoothing, filtering etc.) have to be performed before we can even visually assess the data. These steps will be described here.

First, after loading a run in FSL we ensure ‘Total Volumes’ and TR have been set correctly (these are typically detected automatically), we then set the number of volumes to delete (either 3 for TR of 3, or 5 for TR of 2, deleting 9sec or 10sec of data respectively). As discussed previously (see fMRI – Scan Parameters, section 2.2.3), it takes a short while for the magnetisation to reach a steady state, collecting and discarding dummy volumes helps correct for this. The next parameter to set is the high pass filter, which aims to remove noise at low frequencies. Noise in fMRI can be characterised with a  $1/f$  relationship (Zarahn et al., 1997); the aim therefore, is to remove as much low frequency information as possible without degrading the signal of interest. A common heuristic used in this thesis was to set the high pass filter to twice the length of two block cycles (e.g. for an ABAB design with block length 15sec, set high pass filter to  $2*(15+15) = 60$ sec). Naturally this is not possible with event related designs, in such cases the default high pass filter of 100sec was adhered to, as this should remove the bulk of the low frequency noise.

We next set ‘Motion correction’ using FSL’s ‘MCFLIRT’ (Jenkinson et al., 2002); this estimates all motion against the middle volume within the scan run and corrects for it using rigid body transformations. We also add the estimated motion parameters (translation, rotation) as confound covariates, in an attempt to control for any additional effects of motion. Generally, this should increase statistical power, except in cases where for some reason the motion correlates with a variable of interest. This is most likely to occur if the task (typically requiring motor responses) correlates with said variables; highlighting the importance of task/stimulus orthogonality. Note that in cases where large motion spikes were observed, if we chose to keep the run we explicitly modelled those spikes using FSL’s *fsl\_motion\_outliers* tool. This generates a confound matrix that can be used as an additional step to try and model out large or non-linear motion artefacts.

The next step is to apply spatial smoothing. Whilst it may seem unusual to blur activation after great efforts are spent trying to measure it as accurately as possible,

spatial smoothing can actually increase SNR. Data across voxels will have some level of noise, but this noise should be uncorrelated; i.e. averaging across all voxels would make the noise tend towards zero. Therefore, by smoothing we should reduce noise and make the underlying activation easier to detect. However, if the smoothing is too extensive it will also blur that activation, decreasing SNR. The optimal smoothing size would be one that matched the expected size of the activation, but this is generally unknown. As such, we typically used a heuristic of twice voxel size, so for  $2 \times 2 \times 2 \text{mm}^3$  resolution a Gaussian kernel with FWHM of 4.0mm would be used for smoothing.

For registration, we typically do a three-stage process that has been discussed to some extent previously (see Structural Scans – Axial PDs, section 2.3.4). Briefly, we align functional data to an individual's high resolution space (Average T1) by using an Axial PD as an intermediary. All structural images (including fMRI data) will have been previously skull stripped. We typically keep data in high resolution individual space as most analysis is performed on an individual basis (e.g. based upon ROIs). Exceptions to this will be noted in the relevant chapter. Due to the inhomogeneity gradient caused by the 16-channel coil, these registrations do on occasion fail. As such, all registrations were manually checked and (when necessary) corrected before proceeding further.

Moving on to statistical analysis, we first use 'FILM prewhitening'. To explain why, one of the assumptions of the modelling methods used for fMRI is that of uncorrelated noise. Given that most noise in fMRI follows a  $1/f$  relationship (as previously discussed), that assumption is generally not met (i.e. because noise will be periodic, meaning a given time point is more likely to be similar to the previous time point than not). High pass filtering controls for this to some extent, but we can further address it through prewhitening. By this, we mean FSL will attempt to identify and correct for autocorrelation (the tendency for a time series to correlate with itself at different time points) across each voxel. This helps make the underlying statistics valid, whilst also improving efficiency.

Next, we model our explanatory variables (EVs) based upon onset and duration; including one EV for each condition or event of interest. In some cases, additional EVs are included to model out sources of noise (e.g. blink data or task responses).

These EVs are convolved with a double gamma HRF (Glover, 1999), which combines two gamma functions to capture both the rise and consequent undershoot of the hemodynamic response (Figure 2.1, section 2.2.1). The same temporal filtering as applied to the data is applied here to ensure the model matches the fMRI data as closely as possible. We also include the temporal derivative of each EV; these can account for small temporal shifts in the EV which may capture additional variance (e.g. because the hemodynamic response is not necessarily homogeneous across cortex). Once all EVs are defined, we then set contrasts typically to compare each EV against baseline, and also contrasting EVs against each other (exact contrasts will depend upon design and hypotheses).

Finally, we set ‘post-stats’, namely the thresholding approach used to identify significant activation. Typically, we chose cluster correction with default parameters ( $Z > 2.3$ ,  $p < .05$ ), exceptions will be noted in the relevant chapter. This approach is a less-conservative option than voxel-wise correction (which, due to the extremely large number of voxels sampled, can easily miss true positives). Cluster correction works on the assumption that significant activation is likely to occur in clusters, rather than in isolated voxels. As such, significance can be evaluated on the basis of contiguous clusters increasing statistical sensitivity.

The above details analysis at the level of a single run. Once complete, we then analyse data within-participant using fixed-effects analysis (with same cluster correction applied;  $Z > 2.3$ ,  $p < .05$ ). The alternative would be mixed-effects analysis (FLAME1), but fixed-effects allows greater sensitivity. The difference is that mixed-effects models between-session variance, and so is used when analysing data across multiple participants.

#### **2.4.5 Data Analysis II**

The procedures discussed above produce the ‘final data’, which can now be evaluated. Whilst it might seem as though an inordinate number of steps are needed to reach this stage, those steps are all based upon sound methodological principles. As such, as long as those steps are performed correctly they should increase the quality of the data. We now turn to the two predominant means in which this data is evaluated; percent signal change and representational similarity analysis.

Percent signal change (PSC) is often examined first, as it provides a measure of how much a given ROI responded to a given condition when compared to baseline. For block-design experiments, this was easily computed using FSL's Featquery. Featquery calculates PSC by scaling the parameter estimate by the peak height of the regressor, before dividing by mean activity. As discussed earlier, block-designs allow a good estimate of peak activation for a given condition, as the relatively long stimulation periods allow the BOLD response to plateau. Due to this, Featquery's PSC estimate should be reliable. However, for a rapid-event related design, peak regressor height will depend heavily upon the proceeding events; if two events occur quickly in succession the peak will be greater than if those events were spread over longer periods of time (Figure 2.2, section 2.2.1). This means that peak regressor height is no longer an appropriate measure to calculate PSC. To circumvent this, we calculated PSC based upon published guidelines (Mumford, 2007). Specifically, we model a single isolated event convolved with the double gamma HRF, and calculate height from baseline to peak (ignoring undershoot). This provides a scale factor, which we multiply against the parameter estimate, before dividing by the mean activity. This should put the parameter estimates into units of PSC, which we can then average over a given ROI for comparative purposes.

The second main measure we are interested in exploring is 'neural similarity'. In our event-related paradigms, we can extract the *pattern* of activation unique to a specific stimulus. Once we have the activation pattern for a given ROI, we can correlate it against the pattern elicited by a different stimulus type. If the resultant correlation is high, it implies that the selected ROI responds in a similar manner to those two stimuli. By repeating this process across all pairwise combinations of stimuli in a given experiment, we can generate a *neural similarity matrix* that details similarity across the whole stimulus set (higher correlations within a neural similarity matrix will indicate which shapes elicited more similar patterns of neural activity). This approach is sometimes referred to as 'representational similarity analysis' (Kriegeskorte et al., 2008). To interpret the neural similarity matrices, we also typically create predictor variables, these are similarity matrices based upon known physical properties of the original stimuli. If a given predictor correlates with a neural similarity matrix in a given ROI, then it implies that the representation within that ROI may be based in part on the physical property the predictor was generated

from. Essentially, we correlate multiple known representations with an unknown representation, to try and discern the identity of the unknown. By comparing multiple known representations across different ROIs, we aim to build up a picture of how representations change throughout visual cortex.

### 2.4.6 Data Analysis III

Once numbers have been extracted after calculating percent signal change, or performing representational similarity analysis, that data can then be analysed. The exact analysis procedures used for each study will be discussed in the relevant chapters, however there are two recurring corrections that will be briefly discussed here. Namely, the issue of sphericity, and Fisher Z transformations for correlations.

First, sphericity is an assumption of a repeated-measures ANOVA, an analysis we will conduct frequently due to our use of within-subject designs. Essentially, for sphericity to hold we want roughly equal variance at each difference level across the set of conditions (e.g. for 3 conditions, the variance for 1-2 should be similar to 1-3 and 2-3). We can test for this condition using Mauchly's test; a significant result means sphericity has been violated. If so, the degrees of freedom used to assess the F-ratio can be corrected to avoid false positives, either with the Huynh-Feldt or Greenhouse-Geisser correction. We typically use the latter correction, as it is more conservative and therefore a more stringent test of our hypotheses.

The second consideration is a transformation we typically apply to correlations; the Fisher Z transformation (where  $r$  is the Pearson correlation coefficient):

$$\text{Fisher's } Z = \frac{1}{2} \ln \left( \frac{1+r}{1-r} \right), \quad \text{Pearson's } r = \frac{\exp(2Z) - 1}{\exp(2Z) + 1}$$

We apply this transformation as parametric tests (t-tests, ANOVAs) have the assumption of normally distributed data, an assumption Pearson's  $r$  violates. The Fisher Z transformation produces the normally distributed variable  $Z$ , allowing us to average and statistically compare correlations across groups.

## 2.5 LOC Functional Localiser

In this and the following section we will now discuss the process of identifying regions of interest (ROIs). There are two general principles for exploring neural activity across cortex. First is whole-brain analysis. As the name suggests, this aims

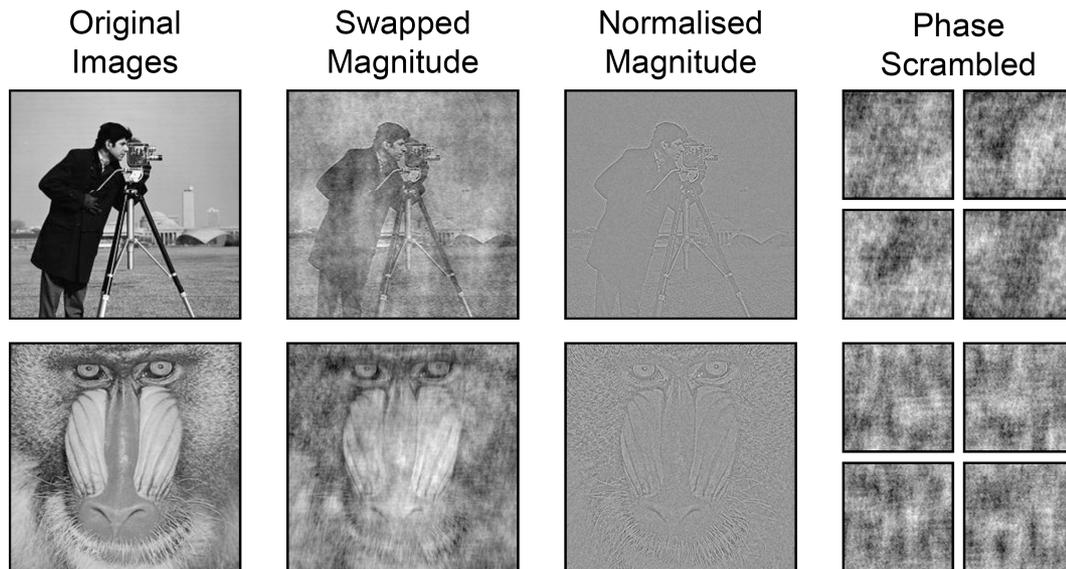
to identify significant activity without *a priori* considerations regarding the location of that activity. Whilst this can be valuable, potentially highlighting activation in regions outside the researcher's expectations, it does come with considerable statistical limitations. Most notably is that of multiple comparisons; due to the sheer number of voxels sampled during fMRI (~ 100,000 or more), the chances of false positives are relatively high without stringent correction. Techniques such as cluster correction may correct for this to some extent, but recent evidence implies this approach may be just as susceptible under certain circumstances (Eklund et al., 2016). Whole-brain analysis can also make results more difficult to compare across studies, as it is hard to determine whether a cluster identified in one study is comparable to that identified in another. Whilst alignment to a standard template (e.g. MNI-152 T1 2mm space) can help to some extent, this still provides relatively poor anatomical specificity due to the nature of the transforms required to perform the alignment.

The alternative (and we would argue preferable) approach is to use ROI based analysis. There are a number of different ways to define ROIs, but typically we run an independent scan specifically designed to exploit known properties of visual cortex. This can either be functional specificity as is the case with functional localisers, or a mapping of the visual world as is the case with retinotopic ROIs. Note that the notion of independence is critically important here, to avoid the issue of circular analysis ('double dipping'); i.e. using the same data set both for ROI definition and data analysis (Kriegeskorte et al., 2009). Otherwise, false positive rates can be greatly inflated. Once defined (a process to be described shortly), these ROIs can be used to compare activity across individuals (and studies) in corresponding regions of the brain. They also reduce the number of comparisons needed, making false positives less likely. However, we should note that this approach does assume the ROI responds homogeneously (at least for percent signal change purposes, less so for similarity analysis). If this is not the case, for example if half the ROI responds whilst half does not, true positives may be missed. This is particularly important for retinotopic ROIs, as we typically only stimulate a small portion of the visual field meaning large parts of the ROI will go un-stimulated. As such, further restriction of ROIs may be warranted if inhomogeneous activation is suspected.

### 2.5.1 Defining the Lateral Occipital Complex (LOC)

The first ROI to explore is the LOC, which has already been discussed in some depth in the previous introductory chapter. Briefly, the LOC is a region (or more specifically, set of regions comprising posterior/dorsal LO and anterior/ventral pFs) that preferentially responds to objects or shape stimuli, over stimuli without coherent form (Malach et al., 1995; Grill-Spector et al., 2001). As such, it is typically identified using an ‘objects vs. scrambled objects’ localiser; the design considerations for this localiser will be addressed here.

Perhaps the most important design consideration is the nature of the ‘scrambled’ control. In Malach et al. (1995), phase-scrambled images were used, and this is still a common approach in the literature. To phase scramble an image, you perform a 2D Fourier transform, splitting the result into magnitude and phase components. The phase component is then randomised, before an inverse Fourier transform is applied to recreate the image. The theory is that this new phase-scrambled image contains the same magnitude spectra as the original, despite being completely unrecognisable. However, using phase-scrambled images has its drawbacks. Firstly, consider an ideal object localiser; this should contrast object stimuli against stimuli that are completely matched with those objects in all visual properties (luminance, contrast, contour information etc.), whilst not actually being objects. Naturally this is impossible (the only perfectly matched stimuli for an image would be that same image), but it gives a metric against which to judge candidate localisers. Returning to the Fourier scrambling paradigm, it is relatively easy to demonstrate that it is the phase spectrum that contains the most salient information about an image, not the magnitude spectrum. For example, you can swap the magnitude (but not phase) spectrums of two images and both remain recognisable, alternatively you can take the inverse Fourier transform of the *just* the phase spectrum (by normalising the magnitude spectrum) to see that it is still recognisable, the same is not true of the magnitude spectrum (see Figure 2.4 for examples). This illustrates that by scrambling the phase spectrum of an image, you remove (or at least greatly alter) the very low-level features that you are trying to control for in the first place. You may get an image that is matched for magnitude, however if the magnitude spectrum contains little salient information then it is hard to argue that these stimuli are generally well matched or controlled for.



**Figure 2.4.** Fourier phase scrambling. The main intent of this image is to demonstrate that after a Fourier transform, it is the phase (and not magnitude) spectrum that carries the bulk of the perceptually salient information. Specifically, we took two images (1<sup>st</sup> column) and performed a 2D Fourier transform on them, splitting the result into the magnitude and phase components. We then either swapped magnitude components across the two images (2<sup>nd</sup> column) or divided the magnitude spectrum within each image by itself (element-by-element) to ‘normalise it’ (3<sup>rd</sup> column). An inverse Fourier transform was then performed to see the result (with some rescaling to 0-255 visible range). Note that in both cases, the images remain readily identifiable, suggesting that the magnitude spectrum contains little salient information, and by scrambling phase information you would lose critical parts of the image. Indeed, the four phase scrambled example images (4<sup>th</sup> column) bear little resemblance to the originals (except with perhaps some coarse low-frequency information preserved).

A frequently used alternative is to just scramble the object images themselves. The technique varies slightly from paper to paper, but generally the image is split into a grid of squares, these squares are then permuted to create an image that contains all the details of the original without giving a holistic impression of an object. Earlier studies using such stimuli showed that LOC’s activity diminishes as the amount of scrambling increases, whereas the activity of lower-level visual areas (e.g. V1-V3) either remains constant or increases slightly (Grill-Spector et al., 1998b; Lerner et al., 2001). For example, Grill-Spector et al. found that using a 4x4 grid (i.e. 16 squares) only slightly diminished LOC’s activity, but an 8x8 grid (64 squares) resulted in a significant reduction. This indicates that by balancing degree of

scrambling against corresponding LOC activity, you can create stimuli that still contain perceptually salient information without substantially reducing power for a contrast. Reinforcing this, participants were still above chance at determining gender in an 8x8 scrambled face image (Lerner et al., 2001), demonstrating that such images are not devoid of meaning.

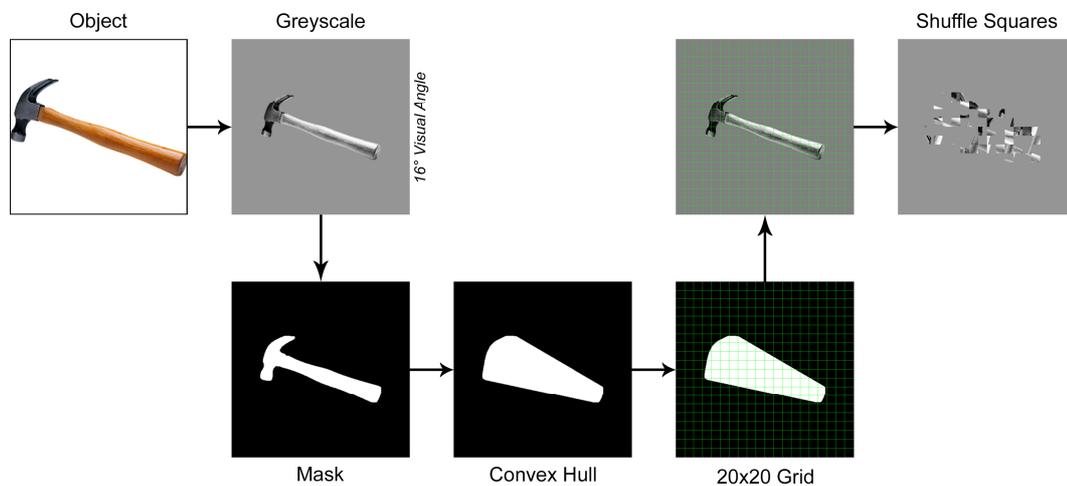
A potential issue with these stimuli is that the scrambling technique introduces ‘lines’ due to the contrast between neighbouring squares. However, both the studies cited above used low-pass filtering on the image, which reduces the salience of those lines. Furthermore, Grill-Spector et al. (1998b) explicitly tested whether this could be a confound by overlying gridlines on top of the object stimuli, finding it had little effect. As such, it seems unlikely that this would be the factor underlying the diminished LOC activity. All in all, these scrambled images activate lower-level visual areas at least as well as object images, suggesting they have similar levels of visual complexity (edges/curves etc.), they retain some perceptual salience, and they are easy to produce. They are not completely matched to object images; contours have to be broken up, and the scrambling technique gives the images a ‘blocky’ appearance, but generally they seem like a better control for contrasts than Fourier phase scrambling.

### 2.5.2 Stimulus Generation

Based upon the above, we chose to ‘block scramble’ our images, rather than phase scramble them (Figure 2.5). First, we acquired 225 PNG (lossless) images of easily recognisable objects and manually extracted from their original backgrounds. These were converted to greyscale with a flattened (equalised) image histogram. This should minimise sudden luminance or contrast changes across images. The images were also rescaled based upon area, such that they subtended  $4 \times 4^\circ$  visual angle<sup>2</sup> on average. We chose to match images on area as aspect ratio could differ considerably depending upon the object, making height or width less appropriate controls. All images were finally placed upon a mid-grey background sized  $16 \times 16^\circ$  visual angle<sup>2</sup>.

To create scrambled stimuli, we split the objects and background into a grid with 20 rows and columns (yielding a square size of  $0.8 \times 0.8^\circ$  visual angle<sup>2</sup>). This grid size means each square should be on par or slightly larger than average receptive field size across V1 (Dumoulin and Wandell, 2008); desirable given that we are trying to

control for ‘low level’ details. All squares lying within the convex hull of the object were then randomly permuted and rotated. A convex hull is the smallest convex region (no inward ‘dips’) that can completely enclose the surrounded shape. We used this convex hull because otherwise certain shapes (e.g. a gun) would be recognisable from their outer contours alone, even though internally they were completely scrambled. Scrambling within the convex hull means that the scrambled objects would contain all local (‘low-level’) details from the original objects, plus the same coarse outline, but they would not be semantically recognisable. As scrambling introduced sharp contrast edges between permuted squares, we applied a Gaussian filter (SD 1px) to both the objects and scrambled objects.



**Figure 2.5.** Object Scrambling Process. Here we show the processing steps used to take a given image of an object, and generate its scrambled counterpart. The exact methodology will be discussed in the main text, briefly we segment the object and background into a grid of squares, then every square lying within the convex hull of that object is randomly permuted and rotated.

### 2.5.3 Stimulus Presentation

The next consideration is how long the above stimuli need to be presented in order to reliably identify the LOC. Essentially, we need to ensure that the region localised in a given person at time one is broadly equivalent to the region localised in that same person at time two. This issue was explored by Duncan & Devlin (2011). They took four participants who underwent ten functional localiser scans over a period of two months (five sessions of two 8min scans, one-back task). The combination of all runs created a ‘gold standard’ ROI, against which they tested multiple analysis

permutations (90 per participant) to create test ROIs using varying amounts of data. These test ROIs could then be compared to the gold standard ROI, to see how many runs are necessary to approximate it. Intriguingly, they found that whilst increasing the amount of data did cause a corresponding increase in degree of overlap between test and gold standard ROIs, this increase peaked and then subsequent data actually impaired overlap. They reasoned that this was because the inclusion of additional data, without controlling for inter-run variance, resulted in more voxels passing threshold without necessarily increasing the overlap with the gold standard ROI. Overall, the most successful performance was obtained either with relatively little data, and a lenient statistical threshold ( $Z > 2.3$ ), or large amounts of data and a much more stringent threshold ( $Z > 5.0$ ). Both options resulted in  $\approx 80\%$  overlap between the test and gold standard ROIs.

We erred on the side of caution, and acquired relatively large amounts of data per participant to localise the LOC. We reasoned that it was better to have the choice of statistical threshold stringency, rather than be forced to use more lenient thresholds through lack of statistical power. Specifically, to localise the LOC in a given participant we ran three 8min localiser scans (TR = 3000ms, TE = 30ms, voxel size =  $2 \times 2 \times 2 \text{mm}^3$ , flip angle =  $90^\circ$ , matrix size  $96 \times 96 \times 39$ , FOV = 192mm). We sacrificed a slower TR for smaller voxel size as a block design was used, meaning we do not necessarily need fast sample rates to identify peak BOLD response. However, the additional spatial resolution of a small voxel size is highly desirable, as it should allow us to identify ROI boundaries with greater precision.

Each localiser run followed an ABAB block design contrasting the object images against scrambled objects. We chose not to use a baseline null period (e.g. a condition with only a grey screen) as theoretically the LOC should respond far less to our scrambled images, thereby allowing it to return to a far lower level of activity during those blocks. Furthermore, by removing null periods we can double the amount of data collected per scan. In total, sixteen object blocks and sixteen scrambled object blocks were used per scan (15sec blocks) with one image presented per second (0.8sec presentation, 0.2sec inter-stimulus interval). To ensure the stimuli were attended to, participants maintained fixation on a central red cross whilst performing a one back task in which there could be one, two or no repeats within a given block. This variation ensures that participants generally have to attend

throughout a block to identify repeats. If, for example, there was one repeat per block and that repeat happened at the start, the participant could potentially lose focus for the remainder of the block. All stimuli were presented centrally against a full screen background set to mid-grey ( $\sim 200\text{cdm}^{-2}$ ).

#### **2.5.4 Data Analysis**

Data analysis for the localiser largely followed the principles set out in the previous section (see Experimental Designs – Data Analysis I, section 2.4.4). Briefly, at the first (individual) level we remove the first three volumes and use a high-pass filter cut-off point of 60sec to correct for low-frequency drift. Motion is corrected for, and motion parameters are added as confound covariates. Spatial smoothing is performed with a Gaussian kernel of FWHM 4mm and FILM prewhitening is used. A single EV is set up to model ‘object’ blocks, with a corresponding contrast of ‘objects > baseline’ (baseline here being scrambled objects). We then run fixed-effects analysis to combine runs within each participant. Example data can be seen later in Figure 2.8.

Once this processing has been performed, it should theoretically be quite simple to then define the regions of interest. However, a principled, objective way to do so proved elusive. As such, the exact methodology used to define the LOC does differ to some extent throughout this thesis. The exact methodologies used will be described in the relevant chapters, however the three core approaches will be detailed briefly here.

The first, and simplest approach was simply to set a relatively stringent cluster (or voxel-wise) threshold, taking all significant voxels as the LOC. Indeed, this is one of the most common approaches in the literature (Duncan and Devlin, 2011). Once voxels are selected, the LOC can then be manually segmented into its posterior/dorsal and anterior/ventral components (LO, pFs respectively). This was typically performed in the volume view of FSL’s FSLView. The reason we chose to move away from this method is that we found the number of active voxels could vary considerably across participants. This is an inherent problem with a functional localiser as opposed to techniques such as retinotopic mapping; there are no objective, concrete boundaries to abide by. Therefore, the size of the resultant ROI depends considerably upon the thresholding applied. Whilst some do apply different

thresholds for participants with more limited activation (e.g. Op de Beeck et al., 2008b), we prefer not to take that approach resulting in quite variable LOC sizes. The second difficulty is that we often observed no discrete split between LO and pFs ROIs. This meant a subjective judgment had to be made regarding where exactly the split should occur; an undesirable approach in terms of replicability.

Our second approach aimed to remove some of the subjectivity by employing a technique based upon averaging (Julian et al., 2012). Briefly, we transform all participants' activity in a given study into standard space (MNI-152 T1 2mm), before thresholding at some value to find voxels for which most participants show activation. This activity is then segmented into LO and pFs ROI *masks*. By transforming the masks back into individual space for each participant, we can identify the participant's peak voxel within those masks and then take a certain number of surrounding voxels as the final ROI (i.e. sphere about peak voxel approach). This technique does have a number of advantages. First, even though it is still subjective, the LO/pFs boundary only has to be identified once in standard space, reducing the chances of error. This also ensures that LO and pFs are defined in roughly the same anatomical locations across participants. Second, the sphere-based approach means that all ROIs share roughly the same number of voxels, accounting for individual variability (plus, this is also a desirable property for representational similarity analysis). The only reason we chose to move away from this method in later work was ultimately based upon pragmatic terms. As this technique relies on averaging across multiple participants, if the same participant is involved in two discrete studies then their final LO and pFs ROIs might actually differ across studies to some extent. Given that a primary aim of ROI analysis is to allow cross-study comparisons, this is naturally undesirable. An alternative would be to slowly build up average LO/pFs masks across all studies (akin to a standard structural image such as the MNI-152 T1 2mm brain), but that would likely still cause shifting ROI definitions until stability was reached. We should also note that this method is not as straightforward to convey as more common alternatives, and it did cause misunderstandings when a recent paper was under review.

For our final approach, we returned to drawing ROIs manually, but this time we drew them when rendered on a partially-inflated surface restricted to grey-matter only (as opposed to drawing in the volume view of FSL's FSLView). This approach

is closest to that used when defining retinotopic ROIs (see next section; 2.6), increasing comparability between the resultant ROIs (e.g. retinotopic and functional ROIs are now both constrained to grey-matter). Drawing on a surface also allows for more discretion, as we can compare the activation to anatomy and other ROIs of interest. For example, we know in most participants that the LOC typically lies near to or overlaps with retinotopically-defined LO-2 (Larsson and Heeger, 2006; Sayres and Grill-Spector, 2008; Silson et al., 2016), so clusters of activation lying well away from this region could be suspect. Finally, this method does have the (debatable) advantage in that it is analogous to that most commonly used in the literature, and is therefore uncontroversial.

We should note that despite using varying approaches to define LOC across studies, we still believe the findings are comparable. Indeed, the variability here simply reflects that of the literature as a whole. The three approaches will largely identify similar clusters of activation; the drive to test different approaches is more being governed by the desire to make that identification as objective and replicable a process as possible.

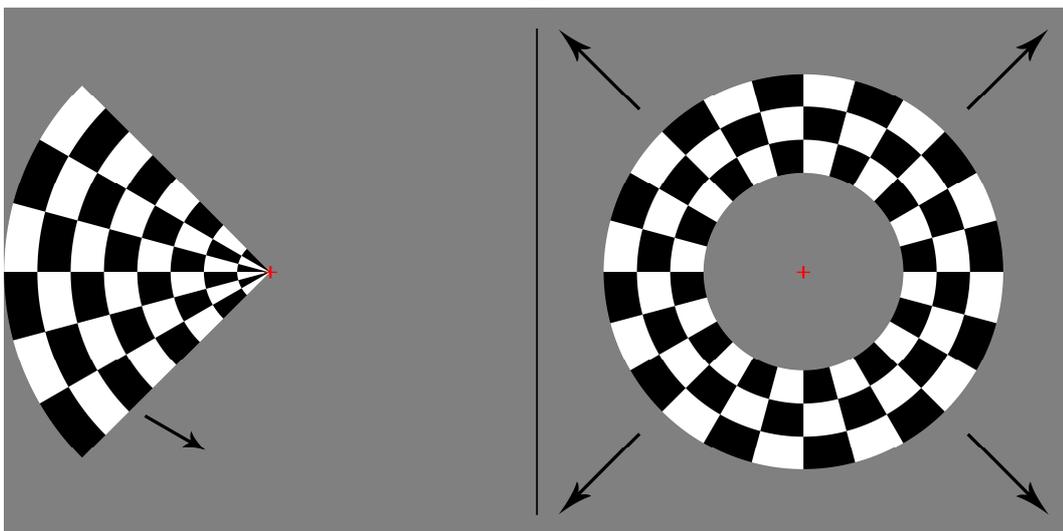
## **2.6 Retinotopic Mapping**

### **2.6.1 Stimulus Generation**

We now move from defining ROIs based upon functional activity, to defining ROIs based upon retinotopy. Retinotopic mapping exploits the fact that early visual areas contain explicit maps of the visual world (Engel et al., 1994; Sereno et al., 1995; Wandell et al., 2007). As such, the aim of retinotopic stimuli is to activate visual cortex as strongly and uniformly as possible. By this, we mean that the stimulus chosen should ideally activate early maps like V1 as well as it activates later maps like LO-2 (although this is not always possible in practice). The stimulus chosen should also be relatively low-level, that is it should consist of quite basic visual properties. To use an extreme example of why this is the case, if just the left side of a face was used for retinotopic mapping, it is possible that in certain areas feedback would ‘fill in’ the right side; making the corresponding spatial location of the activation uncertain. Such ‘filling in’ has been noted as early as V1 (Morgan et al., 2016; note, at time of writing work not peer-reviewed). By using basic stimuli, you

should minimise the risk of participant expectations or visual heuristics (e.g. contour completion) eliciting activity when or where you do not expect any.

Due to these considerations, the stimuli typically used for retinotopic mapping are checkerboard patterns that reverse contrast at some predetermined frequency. These incorporate high-contrast edge and corner information at various orientations, as well as constant luminance changes and so they can elicit robust activation across visual cortex. Throughout our experiments a high contrast (>98%, 400cdm<sup>-2</sup>) radial (circular) checkerboard with a diameter of 21.75° visual angle (filling the projector screen vertically) was used for retinotopic mapping. The checkerboard was split across the diameter into 16 sections, and split about the circumference into 24 segments. Half the segments were black (RGB 0, 0, 0) and half white (RGB 255, 255, 255). Adjacent segments had opposite colours.



**Figure 2.6.** Retinotopic Mapping Stimuli. To map polar angle, we use a 90° wedge-shaped checkerboard that rotates counter-clockwise about fixation. For eccentricity, we use an annulus (or ring) shaped checkerboard that expands from fixation.

### 2.6.2 Stimulus Presentation

As our aim is to systematically map the visual world, we masked the checkerboard with two discrete shapes; a wedge and an annulus (ring), shown in Figure 2.6. The wedge is 90° in size (i.e. a quarterfield, 6 segments) and starts at the upper vertical meridian, before rotating by one segment at a time counter-clockwise about a central red fixation cross. This allows us to map polar angle (e.g. vertical versus horizontal orientations; theta in polar coordinates). In total, a wedge rotates eight times in a

given scan, with duration of 36sec per cycle. This cycle rate should allow the hemodynamic response to return to baseline between successive stimulations, even for ROIs with large receptive field sizes. We typically collected six wedge scans on each participant. In addition to mapping polar angle, we also aimed to map eccentricity; how far into the periphery (from the fovea) something is (rho in polar coordinates). To do this, we used the annulus or 'ring' stimulus. The ring started at fixation with three sections visible (sections 1,2,3), before expanding one section at a time (i.e. next would be sections 2,3,4, then 3,4,5 etc.). The ring wrapped around, such that three sections were always present on screen (i.e. it would transition from sections 6,7,8 to sections 7,8,1). Again, eight cycles each of duration 36sec were used, and we typically acquired two ring scans for each participant. Both wedges and rings were set against a mid-grey screen and they reversed contrast at a rate of 6Hz. The wedge and ring scans both used the same parameters (TR = 3000ms, TE = 30ms, voxel size =  $2 \times 2 \times 2 \text{mm}^3$ , flip angle =  $90^\circ$ , matrix size  $96 \times 96 \times 39$ , FOV = 192mm). We prioritised a smaller voxel size over short TR as we are aiming to localise the boundaries of our ROIs as precisely as possible. Three dummy volumes were collected and discarded at the start of each scan to ensure stable magnetisation was reached.

### **2.6.3 Stimulus Presentation – Future Considerations**

There are a number of design decisions in our retinotopy paradigm that have remained unchanged for the sake of consistency throughout this thesis, but may warrant some examination. The first is the use of task; during our scans participants are just asked to maintain fixation to stabilise eye movements, however such passive viewing may reduce attention and increase fatigue. There is some evidence that encouraging participants to attend to the wedge with a target detection task (i.e. target appears in wedge) whilst maintaining fixation improves map coherence compared to a similarly difficult task at fixation (Bressler and Silver, 2010). However, that study used experienced participants and this is not always feasible in participant recruitment. As such, there is a danger that participants would move fixation towards the attended area negating any benefits that the attention would yield. Nevertheless, the suggested benefits were particularly pronounced in Lateral Occipital areas LO-1 and LO-2, so if the importance of maintaining fixation was stressed then this may prove worthwhile.

The second consideration regards our stimulus size. We used 90° wedges throughout, partially on the basis that these should better activate regions with large receptive fields such as LO-1 and LO-2 (Dumoulin and Wandell, 2008; Amano et al., 2009) because a greater proportion of that region will be activated at any one time. However, there is also a danger that responses will be saturated, indeed some have argued for thinner stimuli when assessing areas with large receptive field sizes (Tootell et al., 1997). To use an extreme example, if the wedge was a full hemifield, it is likely activation in areas like LO-1/2 would never reach baseline after initial stimulation; subsequent stimulations would be occurring too quickly reducing SNR (this issue is analogous to that of TR choice discussed earlier in section 2.2.3). As such, a balance is needed between a stimulus large enough to sufficiently activate the region of interest whilst also being small enough to allow a full return to baseline. In the original LO-1/LO-2 paper, Larsson & Heeger (2006) used a 45° wedge for this very reason. Naturally this means smaller wedges can still be used to identify these ROIs and so the alternative size may be worth trialling.

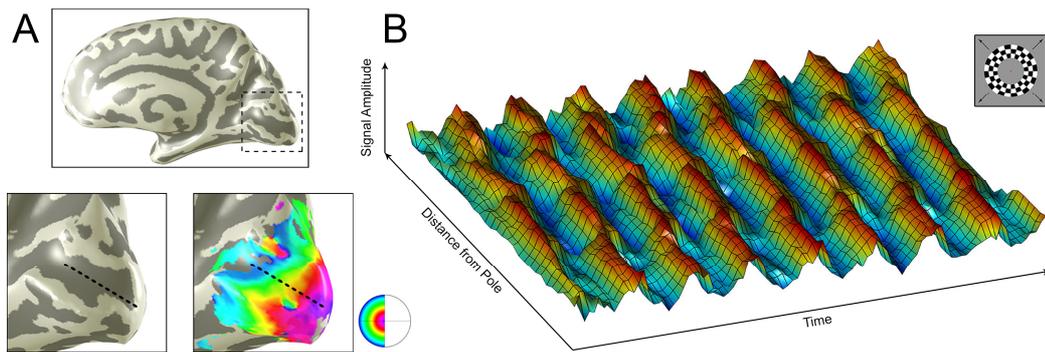
More speculatively, it is possible that smooth rotation, rather than rotation in one-segment increments, may be preferable. Retinotopy relies on a travelling wave of activation, but with incremental movement a given receptive field may jump from un-stimulated to stimulated in one transition. Smoother motion could make this a more gradual process, perhaps making polar angle boundaries more precise. The final consideration in regards to our stimuli is that the ring sections are all of constant size. We know that visual acuity is far greater in the fovea than periphery, and this corresponds to cortical magnification (Duncan and Boynton, 2003). As such, smaller ring sections in the centre with increasing section size as a function of eccentricity would likely better map eccentricity preferences across visual cortex.

Our last consideration is that our wedge stimulus always rotates counter-clockwise. However, due to the slow nature of the hemodynamic response there is some phase lag associated with activation. This can be accounted for to some extent by combining both clockwise and counter-clockwise rotated wedges, providing better phase estimates (e.g. Larsson and Heeger, 2006), and so both scan types are likely worth including.

We should note that despite these design considerations we were able to retinotopically identify all target ROIs in the vast majority of participants with the current paradigm. Nevertheless, interpretation of our data is critically dependent upon accurate ROIs, and so any improvements in retinotopic mapping procedures would be welcome.

#### **2.6.4 Data Analysis**

As mentioned above, retinotopy relies on a travelling wave of activation. As such, analysis of retinotopic data is considerably different from data generated using the block and event-related paradigms discussed above. First, the notion of a ‘travelling wave’ relies on the periodic nature of retinotopic stimuli. Neurons in visual cortex have receptive fields, and the properties of those receptive fields describe the region of visual space that will elicit responses. Those properties can be split into a locus of activation ( $x, y$ ) and measure of spread ( $\sigma$ ) (Dumoulin and Wandell, 2008; Amano et al., 2009). Critically, that locus of activation varies systematically across neurons within a given retinotopic map. Therefore, as a wedge rotates it will periodically activate discrete populations of neurons as it falls into their receptive fields. During our scans, a wedge rotates eight times about fixation; this means a given neuron’s activation will periodically rise and fall eight times in a wave-like pattern (for an example, see Figure 2.7). Crucially, this wave has an associated phase shift dependent upon that neuron’s receptive field properties (namely its polar angle bias). Through the Fourier transform of that neuron’s time series (or cross-correlation), the phase at eight cycles can be extracted. Repeating this across visual cortex then provides a measure of polar angle tuning within each voxel (analogous methods are used to extract eccentricity tuning from ring data). For example, a phase of 0 implies that the neuron’s activity starts rising from the very start of the scan. Given that our wedge starts at the upper vertical meridian, it implies that the neuron is tuned to stimuli situated on that meridian.



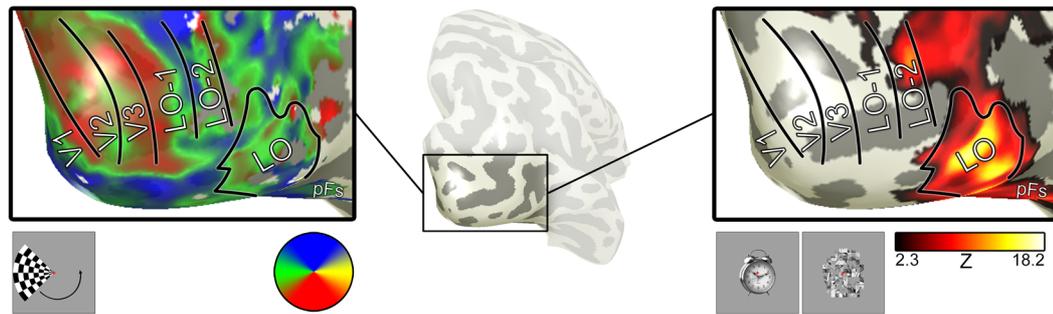
**Figure 2.7.** Travelling Wave Analysis. To demonstrate travelling wave analysis, we took ring (eccentricity) data from one participant, rendered on a surface view in A. The colour gradient reflects differing eccentricity preferences across the cortical surface, with foveal tunings at the Occipital Pole and more peripheral tunings as you move anteriorly. We then sampled voxels across the Calcarine sulcus (i.e. across V1; dashed line in A), and plotted signal amplitude across time for each voxel in B. First, note that across time there are eight peaks, representing the eight cycles in our stimulus presentation sequence. Furthermore, as distance from pole increases, those peaks become shifted in time representing the different eccentricity preferences across cortex. The colour gradient shown in A is generated by mapping the phase of those peaks.

To actually perform the analysis described above we used a software package originating from Stanford, termed MrVista (<https://web.stanford.edu/group/vista/cgi-bin/wiki/index.php/MrVista>). The initial analysis is much the same as that described for block or event-related paradigms. First, for every scan we discard the three dummy volumes, then correct for motion both between and across scans. All data is then aligned to the high-resolution T1-weighted image for that individual. To reduce noise, we collapse (average) all wedge scans into one, then run a correlation analysis to identify the phase for each voxel that best explains its response pattern. An identical process is performed for ring scans. The set of phases across voxels is then restricted to grey matter only, using the grey/white segmentation described earlier (Structural Scans – Brain Segmentation, section 2.3.3), before being mapped in false colour for visualisation purposes. The phase values are also restricted based upon coherence (typically set to 0.20), to reduce the influence of noise. Finally, we create a partially inflated surface view from that same segmentation using ‘MrMesh’ (another part of the MrVista software package), and overlay the phase values on that surface.

### 2.6.5 Identifying Regions of Interest (ROIs)

The final (and most important) step in the retinotopic mapping process is to actually identify our regions of interest. This is performed manually, by ‘drawing’ boundaries on the surface of the brain. We do this based upon the well-documented visual field maps that exist across visual cortex (Engel et al., 1994; Sereno et al., 1995; Larsson and Heeger, 2006; Wandell et al., 2007). Specifically, we typically aim to identify V1, V2, V3, V4 (sometimes termed hV4 to distinguish it from Macaque V4), LO-1 and LO-2 in both hemispheres of each participant (Figure 2.8). To discriminate between said ROIs, we look for phase-reversals at the horizontal or vertical meridians; these represent boundaries between discrete regions. To be confident in the ROI, we also try to ensure that there is a smooth transition in phase values across its surface, however on occasion noise must be accounted for. Our target ROIs have the following properties (note that all described maps are of the contralateral visual field):

- **V1:** A full hemifield representation, moving ventrally-to-dorsally transitions from the upper vertical meridian (UVM) to the lower vertical meridian (LVM) (i.e. a flipped representation of the visual field).
- **V2:** Split into two quarterfield representations; V2v and V2d (ventral and dorsal V2 respectively). V2v lies adjacent to ventral V1, extending from the UVM to the horizontal meridian (HM). V2d lies adjacent to dorsal V1, extending from the LVM to the horizontal meridian (HM).
- **V3:** Again, split into two quarterfield representations; V3v and V3d (ventral and dorsal V3 respectively). V3v lies adjacent to and is the mirror reversal of V2v (i.e. HM-to-UVM). V3d lies adjacent to and is the mirror reversal of V2d (i.e. HM-to-LVM).
- **V4:** A full hemifield representation, adjacent to V3v. Extends from the UVM to the LVM.
- **LO-1:** A full hemifield representation, adjacent to V3d. Extends from the LVM to the UVM. This is sometimes said to form the ventral part of a ‘Y’ shape through the boundaries of V3d, LO-1 and V3A/B (an additional pair of ROIs which we do not typically aim to identify).
- **LO-2:** A full hemifield representation, adjacent to and mirror of LO-1 (i.e. UVM-to-LVM).



**Figure 2.8.** Example Retinotopic Maps. Here, we have taken the inflated surface of one participant's right hemisphere and rendered polar angle preferences in false colour across the cortex. This demonstrates the transitions between the hemifield map V1, quarterfield maps V2d and V3d then hemifield maps LO-1 and LO-2. Boundaries between adjacent ROIs are drawn across horizontal or vertical meridians. Analogous boundaries are observed between ROIs on the ventral surface (not shown). For illustrative purposes, we have also included activation elicited by an 'Objects > Scrambled' (LOC) localiser, LO itself being defined based upon a sphere around peak voxel (see Chapter 5, section 5.3.2). Note that the activity falls largely in regions undefined by retinotopic maps, with some slight overlap observed between LO-1 and (to a greater extent) LO-2.

The ROI drawing process follows the above sequence (V1 to V4, then LO-1 and LO-2); by drawing each in turn we can ensure that the shared boundaries are allocated as accurately as possible. However, because these ROIs are drawn by hand, the final boundaries sometimes overlap. To address this, we run a custom-written script that removes ROI intersection. Specially, for a given pair of ROIs (e.g. V1, V2d) it identifies all shared voxels (if any). It then calculates the distance between each of those voxels and the two ROI's centroids, turning those distances into Z-Scores. To do this, we take all voxels in (e.g.) V1 and calculate the average distance plus standard deviation (SD) from V1's centroid. Then for each intersecting voxel we subtract that mean and divide by the SD. This is repeated for the second ROI (i.e. V2d in this scenario). For every intersecting voxel, we now allocate it to the ROI for which it has the lowest Z-Score. Z-Scores are used instead of standard Euclidean distance because otherwise smaller ROIs (in this case V2d) would naturally have smaller distances to any intersecting voxels. Z-Scores provide a means to predict which ROI is most likely to contain any given voxel. After this process is completed we now have a complete set of ROIs, with which we can conduct analysis.

## 2.7 Radial Frequency (RF) patterns

In the final two sections of this chapter, we will discuss two methods for stimulus creation that are used throughout the thesis. First, we have radial frequency (RF) patterns (Wilkinson et al., 1998).

RF patterns are broadly circular, sinusoidally-modulated outlines that are defined in terms of the number of lobes they have (e.g. a ‘three-lobed’ pattern has a clover-like appearance) and the size of those lobes (Figure 2.9). Specifically, RF patterns are defined in polar coordinates by calculating rho ( $r$ ) as a function of theta ( $\theta$ ) using the formula:

$$r = r_0(1 + A(\sin(\omega\theta + \phi)))$$

Here, theta ( $\theta$ ) represents the angles around a circle’s perimeter, allowing the sinusoidal modulation of that perimeter by altering frequency ( $\omega$ , governing number of lobes) and *amplitude* ( $A$ , governing size of lobes). Rotation can be set by altering phase ( $\phi$ ). The mean radius ( $r_0$ ) governs the average size of the stimulus, typically set in degrees of visual angle.

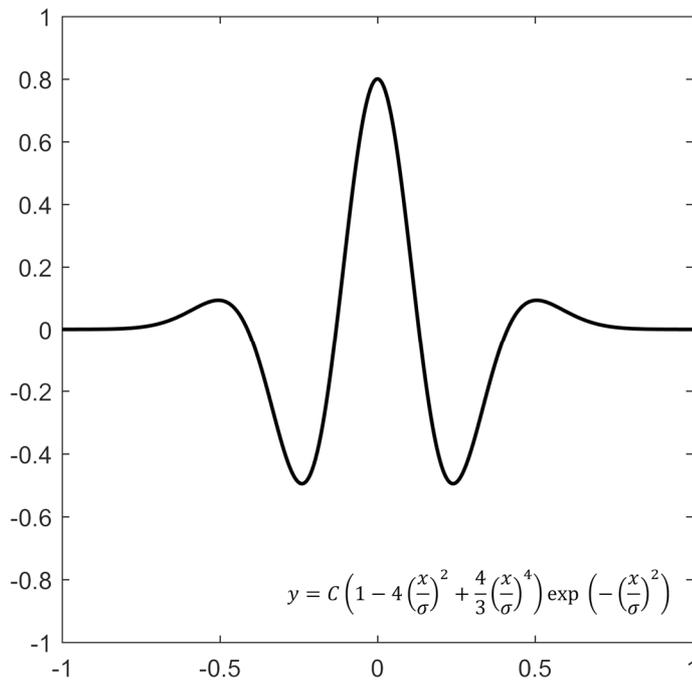


**Figure 2.9.** Example Radial Frequency (RF) Patterns. In the middle we have a three-lobed RF pattern. The RF pattern to the left depicts an increase in *amplitude*; this creates the impression of more pronounced lobes. The RF pattern to the right depicts an increase in *frequency*; namely a switch from 3 to 5 lobes.

RF patterns have a number of properties that make them appealing stimuli. First, being mathematically defined they are easy to parametrically manipulate. Parametric definition is critically important for stimuli, as it means those parameters can be slowly and methodically manipulated to determine exactly what properties of the stimulus are being responded to. Essentially, it provides very tight control over the

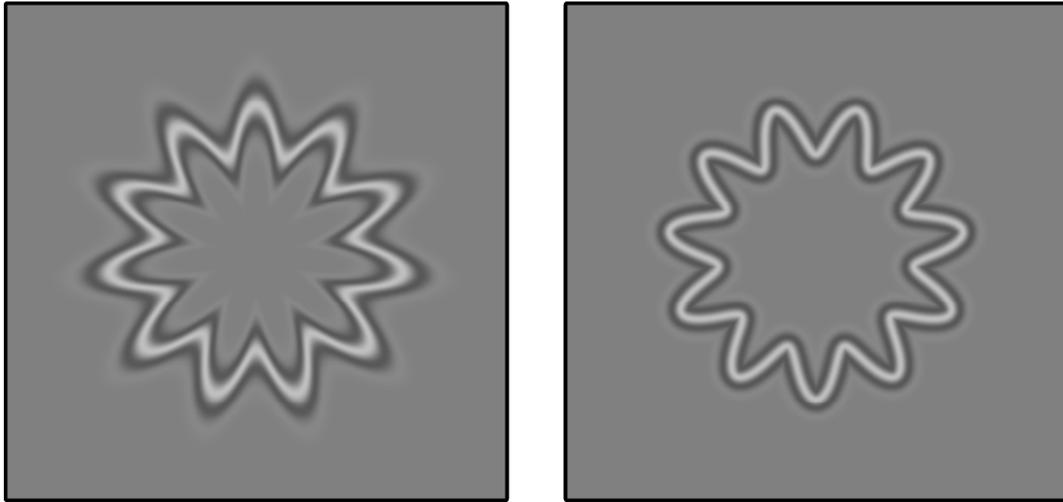
differences between each stimulus in our stimuli set. This also has advantages when aiming to control for low-level visual properties. Being broadly circular, RF patterns also contain considerable amounts of curvature information; a potentially salient property for shape-processing regions (Pasupathy and Connor, 1999, 2001, 2002; Habak et al., 2004; Haushofer et al., 2008a; Kayaert et al., 2011; Yue et al., 2014). Finally, there is evidence that LO-2 is causally involved in the processing of three-lobed RF patterns (Silson et al., 2013), making them a highly relevant stimulus for studying this region's function.

To display the shapes (as in Figure 2.9), the outer contours were rendered against a mid-grey background using the radial fourth derivative of a Gaussian (D4) (Wilkinson et al., 1998), as plotted in Figure 2.10. This has two key advantages. First, this function integrates nearly to zero across its width, therefore it controls for gross luminance changes. Secondly, because the function contains a peak (high luminance) flanked by two dips (low luminance), there is good contrast with which the contour can be detected. This contour also gives the impression of a Gabor element-like appearance. A Gabor element is a patch of sinusoidally-modulated luminance in a Gaussian aperture, and these are ideal for activating simple cells in striate cortex which respond to edges and gratings (Hubel and Wiesel, 1959). Indeed, banks of Gabor filters have been used to model low-level visual responses (termed "GIST descriptors"; Oliva and Torralba, 2001). As such, our stimuli should elicit robust activation in early visual cortex, which will hopefully carry through to later areas of interest. These stimuli are typically rendered at 50% contrast, as this makes them easily detectable and greater contrasts may cause local light adaptation (Kingdom and Whittle, 1996).



**Figure 2.10.** Fourth Derivative of a Gaussian (D4). This is the D4 pattern used to render radial frequency patterns. Here, contrast ( $c$ ) was set to 0.8 and  $\sigma$  (determining peak spatial frequency) was set to 0.25. Note that in polar coordinates,  $x$  would instead be  $(r - r_0)$ .

We should note that we had to adapt the rendering methodology slightly to account for more complex shapes. Typically, the D4 function is used in polar coordinates. For a given pixel, you calculate the radial distance (i.e. the distance along an imaginary line passing through the origin) between that pixel and its nearest contour point. The pixel is then shaded based on the D4 function at that distance. However, deviations from concentricity will cause rapid changes in radial distance, resulting in unnatural shading. This means the approach is inappropriate both for RF patterns with high frequency (e.g. Figure 2.11), RF patterns with high amplitude (deviating away from circularity), or other non-circular shapes. To address this, we instead used the distance transform to shade pixels. Specifically, for a given pixel we simply calculate the shortest distance to the contour along *any* direction, and then shade it with value of the D4 function at that distance. This allows us to render any contour in this manner, not just circular outlines.



**Figure 2.11.** D4 Rendering Method. The left pattern represents an RF pattern (frequency 11) rendered using the fourth derivative of a Gaussian (D4) in polar coordinates. Changes in curvature cause rapid changes in radial distance, resulting in pinching nearer the shape's centroid and spread closer to the perimeter. We solve this by rendering shapes based upon a distance transform (see main text for details), resulting in the more uniform boundary shown on the right.

## 2.8 Fourier Descriptors (FDs)

Whilst radial frequency patterns are undoubtedly useful, they are a relatively artificial stimulus set (poor ecological validity) with limited variability. We aimed to expand our stimulus creation toolkit through the use of Fourier Descriptors (FDs) (Zahn and Roskies, 1972). Briefly, FDs allow you to describe any shape's contour as a series of sine waves, essentially transforming the contour into the frequency domain. Critically, we can manipulate the shape in the frequency domain, before applying the inverse Fourier transform to view the result. This approach allows us to parametrically vary any shape contour, not just those that are founded upon mathematical definition. We calculated the set of FDs for a given image using the following procedure.

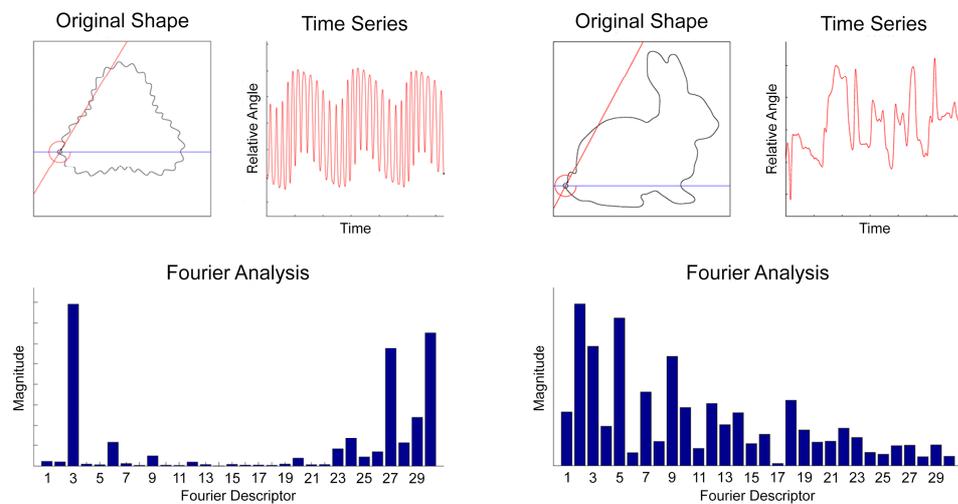
First, the outermost boundary of a given shape was extracted and moving average smoothing was applied, to correct for pixelation. However, as different shapes will have different levels of detail, the smoothing was customised on the following basis. First, we smooth with smoothing window of 1 (i.e. every point is replaced with the average of that point and the points immediately adjacent to it) and calculate the average discrepancy between the original and smoothed boundaries; call this the

initial discrepancy. We then increase the smoothing window until the discrepancy between the original boundary and current smoothed boundary is three times that initial discrepancy. For shapes with lots of high frequency details, smoothing will have drastic effects and so the discrepancy should rise sharply as a function of smoothing window. Conversely, in shapes with broader outlines a greater degree of smoothing can be applied before the underlying contour is corrupted. Essentially, this means that smoothing was tailored for each shape to maximise smoothing whilst minimizing loss of high frequency details. Finally, the resultant smoothed boundary was then interpolated using a periodic cubic spline curve to 4,096 points (whilst probably not necessary, the Fourier transform is more efficient with  $2^n$  points).

Next, we need to essentially turn the contour into a time series so that we can perform Fourier analysis upon it (time here conceptually represents the ‘time’ taken to traverse the complete contour from an arbitrary starting point, thus we plot how that contour changes over time). There are a number of ways to do this, however we had the constraint that the method chosen must be reversible. This rules out certain approaches, for example, a commonly used method is to track about the contour and plot the distance along the radial line from the current contour point to the shape’s centroid. This method is advantageous as it is quite resilient to noise about the contour. However, if the shape doubled back upon itself at any point (e.g. as on an umbrella handle) then this information would be lost, as the orientation of the radial line is not recorded. Due to this, we instead chose to model orientation changes as a function of distance about the contour.

Specifically, for every point around the contour, we calculated both the distance and angle to the next point. The distance was then normalised to the range  $0-2\pi$ , and we also removed the linear trend in the angles. To explain the latter step, imagine traversing about the contour of a circle; the angle from one point to the next would increase linearly as a function of distance. By removing this linear trend, a circle would now be represented as a flat line, which makes some intuitive sense. This also serves the Fourier analysis, as (for example) the angles in a three-lobed RF pattern with linear trend removed would now be represented by  $\text{Sin}(3\theta)$ . Naturally, this can easily be expressed in the frequency domain, demonstrating that FDs can tell us something meaningful about the underlying shape information. The Fourier analysis itself was performed on the set of angles (Figure 2.12). Critically, we save out

enough information to fully recreate a shape from its Fourier spectrum, making this process completely reversible. Therefore, shapes can now be manipulated in the Fourier domain and transformed back to view the results. This comes with one caveat; manipulations in the frequency domain did sometimes result in non-closed boundaries. However, this was easily resolved by removing linear trends from the resultant x- and y-coordinates after performing the inverse Fourier transform and recalculating the shape. The exact uses of this manipulation will be discussed in the relevant chapters



**Figure 2.12.** Fourier Descriptors (FDs). The left panel depicts the process of calculating a set of FDs for an RF3 pattern with added high frequency noise. We first track around the original shape (top left), calculating the angle between successive pairs of points on that shape’s contour. After removing the linear trend inherent when traversing around a shape, these angles can be expressed as a ‘time series’ (top right), where ‘time’ represents the time taken to traverse the complete contour. Finally, we perform a Fourier analysis on that time series, the bottom graph shows the magnitude spectrum of the analysis (with DC component removed). Note the high magnitude at FDs 3, 27 and 30. The first represents the coarse RF3 pattern of the shape, the latter two reflect the high frequency noise at 9 and 10 cycles per lobe respectively. This demonstrates that the FD approach can yield valuable information about the original shape, and it is not simply a mathematical abstraction. In the right panel, we perform exactly the same process on a more complex object; namely a rabbit.

## 2.9 Summary

To summarise this chapter, we have given a broad overview of fMRI, from its underlying principles to the techniques we will employ to utilise it. Specifically, we have covered the preliminary scans that all participants undergo; namely structural scans and scans to identify ROIs (functional localisers, retinotopy). We have also covered the design decisions and analysis procedures we consider for new fMRI experiments. Finally, we have detailed the procedures used to create the stimuli that will be used throughout this thesis. We now move on to empirical chapters.

## **Chapter 3. Responses to Orientation and Shape Processing Tasks in Retinotopic Regions LO-1 and LO-2**

### **3.1 Abstract**

A recent study using TMS (Silson et al., 2013) found a double dissociation between two retinotopic regions in Lateral Occipital Cortex; LO-1 and LO-2. LO-1 was found to play a causal role in performing an orientation task (judging the orientation of a sinusoidal grating) whereas LO-2 played a causal role in a shape task (judgements of curvature). Whilst compelling, the two tasks used in the aforementioned study differed in more dimensions than just orientation and shape. To address this, we ran a block-design fMRI study with two categories; orientation and shape. However, in our experiment the stimuli used across blocks (three-lobed radial frequency patterns) were identical; the key difference was the task participants performed. During orientation blocks, participants made judgements about the relative orientations of two bilaterally presented shapes. During shape blocks, participants made judgements about which of the two shapes was spikier (i.e. curvature judgement). The orientation and shape differences between the bilateral shapes had been set such that individuals were capable of discriminating those differences at equal levels of performance psychophysically, although performance differences did emerge during scanning. We found that LO-2 showed significantly greater activity generally when compared to LO-1, reinforcing its role in shape processing, however it did not dissociate between tasks. In contrast, LO-1 showed a surprising significant preference for the shape processing task. No between-task differences were identified in earlier retinotopic regions V1 to V4. We discuss possible implications of this finding, including potential differences in local versus global image processing, plus how our results may inform theories regarding the role that these two regions play in shape perception.

### 3.2 Introduction

In this first empirical chapter, we aim to explore further the properties of two retinotopic regions that Larsson and Heeger (2006) identified in Lateral Occipital Cortex, termed LO-1 and LO-2. These regions appear to overlap with the object-selective Lateral Occipital Complex (Malach et al., 1995; Grill-Spector et al., 2001; Sayres and Grill-Spector, 2008) to some degree and, as such, understanding the role of these two areas may shed some light on just how our brain performs object recognition. The former area, LO-1, appears to play some role in orientation processing, both for standard grating stimuli plus second order and illusory edge stimuli (Larsson et al., 2006; Montaser-Kouhsari et al., 2007). Conversely, LO-2 appears to be more tuned for shapes and has little orientation sensitivity (Larsson and Heeger, 2006).

A key study demonstrating this dissociation used transcranial magnetic stimulation (TMS) on LO-1 and LO-2, whilst participants performed orientation and shape discrimination tasks (Silson et al., 2013). The orientation task used pairs of sinusoidal gratings presented successively through a circular aperture with diameter  $4^\circ$  visual angle. Participants had to state whether the second test stimulus was more vertical or horizontal than the previous reference stimulus. The shape task used an identical paradigm, comparing three-lobed radial frequency (RF) patterns. The participant's task was to judge whether the test shape was spikier or smoother than the reference shape (i.e. a curvature processing task). To ensure that the tasks could not be solved using local luminance cues, the phase of the sinusoidal gratings was randomised in the orientation task, and the orientation of the RF shapes was randomised in the shape task. Using these tasks, a double dissociation was found; TMS to LO-1 impaired performance on *just* the orientation-processing task, whereas TMS to LO-2 impaired performance on *just* the shape-processing task. This implies that despite their close cortical locations, these two regions are performing at least some independent functions.

Whilst compelling, Silson et al.'s stimuli do differ on more than one dimension, so it is premature to conclude that the dissociation is entirely due to orientation and shape specificity. For example, the orientation task involved attention to internal, local features whereas the shape task necessitated attention to the global profile of the RF pattern's contours. Due to this, the orientation task could potentially be solved with

smaller receptive fields than the shape task and LO-1 has generally smaller receptive fields than LO-2 (Amano et al., 2009). As such, this is a possible explanation for the dissociation identified. There are also more subtle differences, for example the orientation task involved straight lines whereas the shape task involved curvature, furthermore the shape task had closed contours whereas the orientation just had ‘open’ parallel lines. None of these differences undermine the double dissociation identified; they just raise the question of which difference caused that dissociation.

The aim of this study is to explore that question, specifically asking whether an orientation/shape dissociation will still be identified in LO-1 and LO-2 when both tasks involve some form of shape processing. Our shape task will be virtually identical to the one described by Silson et al.; participants will make judgements about the relative spikiness or smoothness of a three-lobed RF pattern. The critical difference lies in the orientation task; in this study participants will judge the relative orientation of an RF pattern, rather than sinusoidal lines. Furthermore, in both tasks the stimuli presented will differ in both orientation and shape simultaneously; the only difference is which manipulation participants will be attending to. This ensures that our stimuli are perfectly matched across tasks, addressing most of the differences that were identified in Silson et al.’s study. This allows us to test whether LO-1 is generally sensitive to orientation information, regardless of the specific information carrier.

Based upon previous results demonstrating higher shape sensitivity in LO-2 (Larsson and Heeger, 2006; Silson et al., 2013), we predict that LO-2 will generally respond more for both our orientation and shape tasks than LO-1, simply due to the presence of a shape. It is also expected that LO-2 will be more active for the shape processing task as performance on such a task was impaired when LO-2 was disrupted via TMS. As TMS to LO-1 had no effect on the shape processing task, we hypothesise that the orientation task will elicit greater activity in this area. Such a finding would also reinforce orientation specificity, and point to a more general form of orientation processing in that region.

### 3.3 Methods

#### 3.3.1 Participants

Eleven participants (6 females, ages 20-26, 10 right-handed, all gave informed consent) were recruited from the University of York's Department of Psychology. Each participant underwent two psychophysics sessions, one high-resolution structural scanning session, one retinotopic-mapping scanning session and one functional MRI scan.

The structural and retinotopic scans were carried out and analysed as described in the Methods chapter (Chapter 2, sections 2.3, 2.6).

#### 3.3.2 Psychophysics

For every participant, we aimed to match performance on both the orientation and shape discrimination tasks. This was achieved psychophysically, using the method of constant stimuli (avoiding habituation effects). The general approach was this; a reference shape was presented for 200ms, then after an inter-stimulus interval of 1,200ms a test shape would appear, again for 200ms. The next trial would then occur either after the participant's response or 1,500ms, depending on which came last.

The stimulus used was a three-lobed radial frequency (RF) pattern (Wilkinson et al., 1998) defined using the formula:

$$r_0(1 + A(\sin(\omega\theta + \phi)))$$

For the reference shape,  $r_0$  (mean radius) was set to  $4^\circ$  of visual angle,  $A$  (radial modulation amplitude) was set to 0.2,  $\omega$  (radial frequency) was set to 3 and  $\phi$  (angular phase) could be  $20^\circ$ ,  $40^\circ$ ,  $60^\circ$ ,  $80^\circ$  or  $100^\circ$ . Test shapes were created by subtracting from or adding to both  $A$  (adjusting amplitude, or smoothness/spikiness) and  $\phi$  (adjusting orientation). The shapes were rendered using the radial fourth derivative of a Gaussian (D4):

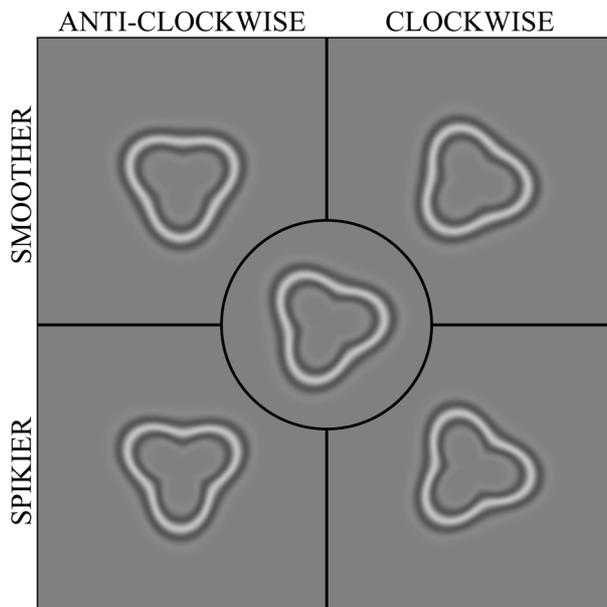
$$D4 = C \left( 1 - 4 \left( \frac{r - r_0}{\sigma} \right)^2 + \frac{4}{3} \left( \frac{r - r_0}{\sigma} \right)^4 \right) \times \exp \left( - \left( \frac{r - r_0}{\sigma} \right)^2 \right)$$

Here,  $C$  (pattern contrast) was set to 0.5 and  $\sigma$  (determining peak spatial frequency) was set to 1. Note that these stimuli were generated before we switched to the

alternative rendering method outlined in the Methods chapter (Chapter 2, section 2.7). However, at threshold the profile itself is unlikely to influence discriminations.

The RF shapes were vertically centred  $10^\circ$  of visual angle into the periphery and placed upon a mid-grey background (RGB 128 128 128), a central red (RGB 255 0 0) fixation cross (0.5 degrees of visual angle) was used to stabilise eye movements. For every participant, we presented the stimuli in the hemifield contralateral to the hemisphere with the clearest Lateral Occipital retinotopic maps.

To assess psychophysical thresholds, participants were seated 57cm away from a CRT monitor in a dark and quiet room. First, they underwent one practice trial with arbitrary, easy conditions. This ensured that participants understood the nature of the manipulations they were to be tested upon. For the orientation task, this meant a change of  $\pm 15^\circ$  (e.g. if the reference shape appeared at  $20^\circ$ , it would be followed by a test shape at  $5^\circ$  or  $35^\circ$ ), for the shape task a value of  $\pm 0.08$  was used (e.g. as the reference shape had an amplitude of 0.20, the test shape would have an amplitude of 0.12 or 0.28). The participant's task was to assess whether the test shape was rotated anticlockwise or clockwise (orientation task), or appeared smoother or spikier (shape task) relative to the reference shape (Figure 3.1). Responses were made via key press; 'Z' for anticlockwise or smoother, 'X' for clockwise or spikier.

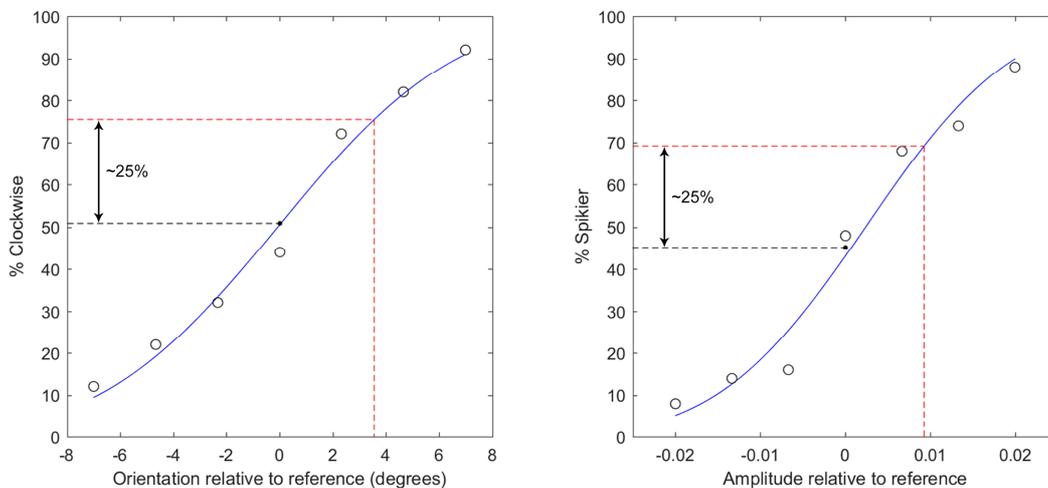


**Figure 3.1.** Example stimuli. An example reference shape is in the centre, and all possible test shape combinations are in the surround. The orientation and shape manipulations here represent three times the detection threshold for one representative participant, i.e. these are the shapes that the selected participant would view during their fMRI scan (see fMRI methods).

After the practice trial, a number of test runs (typically 1-3 runs) were conducted to determine a suitable range over which performance would be tested. We were essentially aiming to find the point at which participants would be almost completely

accurate, without being at ceiling. For example, if a participant could reliably discriminate orientations with a difference of  $9^\circ$  then we would test them at seven linearly spaced intervals with  $\pm 9^\circ$  at the extremes (i.e.  $-9^\circ$ ,  $-6^\circ$ ,  $-3^\circ$ ,  $0^\circ$ ,  $3^\circ$ ,  $6^\circ$ ,  $9^\circ$ ). This ensured a wide range of responses allowing a good psychometric fit.

With the range set, participants performed five runs of the psychophysical test, with 70 trials per run (i.e. 10 trials at each value in their range). As there was not necessarily a correct answer (e.g. at  $0^\circ$  change), we chose the proportion of times that the participant responded clockwise or spikier as our estimate of performance. Their scores across all five runs were averaged and then approximated using the best fitting cumulative Gaussian (sigmoidal) curve, so that we could estimate and save out individual thresholds for the fMRI task. To calculate a threshold, we take the 25% and 75% clockwise or spikier response points from the sigmoid, find the range between them, then add half this range to the participant's mean response to get a threshold. For example, if the participant responded clockwise 50% of the time (as ideally they should), their threshold would be at the 75% point. However, biases one way or another will shift the threshold correspondingly (see Figure 3.2 for an



**Figure 3.2.** Example sigmoid fits for one participant. The black circles represent participant's data, the black dot (and corresponding black dashed line) represents the participants mean response, the blue line represents the sigmoid fit and the red dashed line represents the participant's threshold. In the left graph (orientation judgements), the participant's mean is almost 50%, leading to a threshold of 75%. However, in the right graph (amplitude judgements) the participant shows a slight 'smoother' bias (mean of ~45%), and so the threshold is shifted accordingly. In both cases, the threshold is approximately 25% higher than the participant's mean response.

example). This approach should ensure the threshold value chosen is symmetric about the participant's mean, providing their bias remains stable.

The above procedure was performed twice per participant, first for the orientation task and then for the shape task. As both orientation and shape changed simultaneously in a given run, we had to set the shape manipulation manually when participants were first doing the orientation task (as shape thresholds were yet to be determined). A value of 0.014 was chosen based upon average thresholds from previous work.

### 3.3.3 Functional MRI

All participants underwent three functional MRI runs (16 channel coil; TR = 2000ms, TE = 30ms, voxel size =  $3 \times 3 \times 3 \text{mm}^3$ , flip angle =  $90^\circ$ , matrix size  $64 \times 64 \times 26$ , FOV = 19.2cm). Each run used a block design, with 24 blocks (15sec each) alternating between the orientation and shape tasks (i.e. ABAB design). There were 10sec baseline periods interleaved between blocks during which participants just maintained fixation. A given block was broken down into an initial 0.6sec period where the fixation flashed red (to cue stimulus onset), followed by 8 trials (0.4sec stimulus presentation then 1.4sec response window).

During the scan, participants again performed orientation and shape discrimination tasks, however we decided to move from serial to simultaneous bilateral presentation. Whilst matching psychophysical and fMRI protocols would have been preferable, this modification had two key advantages. First, it essentially doubled the number of trials per run, increasing statistical power. Second, it also meant both visual hemifields would be involved in the orientation and shape processing tasks, allowing the use of LO-1 and LO-2 in both hemispheres as regions of interest. Changing the psychophysical protocol was not possible, as the thresholds were also being used for a separate TMS study in which stimulus presentation had to be constrained to a single hemisphere.

With this new paradigm, a reference shape would be presented on either the left- or right-hand side of the screen ( $10^\circ$  eccentricity from fixation), again at an orientation of  $20^\circ$ ,  $40^\circ$ ,  $60^\circ$ ,  $80^\circ$  or  $100^\circ$ . The test stimulus would simultaneously appear on the other side and it could be smoother or spikier, plus rotated anti-clockwise or clockwise relative to the reference shape. The mean radius of both reference and test

shapes was kept to  $4^\circ$ . The orientation and shape discrepancies were set to three times the thresholds extracted earlier using psychophysics. This was because bilateral presentation made the task considerably harder than the earlier psychophysical protocol, and pilot tests indicated that tripling thresholds made the task feasible without putting performance at ceiling. The participant's task was to determine which side had the most clockwise-rotated shape in the orientation blocks, and which side had the spikiest shape in the shape blocks. Responses were made via a serial response box in their right hand. As stimuli were identical in each task, the fixation point was either an 'o' to cue the orientation task, or a '+' to cue the shape task. A given block of 8 trials contained all permutations of the reference and test stimuli possibilities (reference stimulus being on the left- or right-hand side with the test stimulus being anti-clockwise or clockwise and smoother or spikier), presented in a random order. The reference shape was presented at each of the five possible orientations at least once, plus three orientations twice (again, randomly ordered).

Functional data were analysed using FEAT (fMRI Expert Analysis Tool). For every individual we first ran three analyses, one per scan run. These used a high-pass filter cut-off point of 50sec (length of two blocks plus two baseline periods) to remove low-frequency drift. Motion parameters were estimated using MCFLIRT to correct for head movement during the scan, and these were also entered into the model as confound covariates. Spatial smoothing was set to 6mm (twice voxel size) and FILM prewhitening was used. Scans were first aligned to the average 16-channel T1, and then to standard space (MNI-152 T1 2mm brain). Two contrasts were set up to compare activity during the orientation and shape tasks over baseline.

Next, a second level fixed effects analysis was run for each participant to simply combine their results from the individual runs. ROI analysis was performed at this stage, using FEATquery on the COPE (contrast of parameter estimates) statistics for each task over baseline. COPE values were converted into percent signal change.

## 3.4 Results

### 3.4.1 Behavioural Analysis

From the psychophysics, the final average thresholds were  $5.96^\circ$  (SD  $1.80^\circ$ ) for the orientation task and 0.014 (SD 0.005) for the shape task. The latter result justifies our earlier decision to use 0.014 as the starting amplitude difference during the orientation task (before individual shape thresholds had been determined), and suggests shape performance here is at least comparable to earlier work.

To check participant's performance on the orientation and shape tasks, we ran a  $2 \times 3$  repeated measures ANOVA comparing task with scan run (to look for practice effects). Mauchly's test indicated that the assumption of sphericity had been violated for the main effect of run ( $\chi^2(2) = 7.38, p = .025$ ), therefore degrees of freedom were corrected using the Greenhouse-Geisser estimates of sphericity ( $\epsilon = 0.64$ ). A significant main effect of condition was found ( $F(1,10) = 37.29, p < .001$ ) indicating that participants were substantially better at the shape task. No significant main effect of run was found ( $F(1.28, 12.83) = 3.00, p = .100$ ) and there was no significant interaction ( $F(2,20) = 0.11, p = .898$ ).

As no main effect of run was identified, we collapsed across all runs to check whether participant's average performance was statistically different from chance (50%). Using two-tailed one-sample t-tests we found that this was true for both the orientation task (Mean 65.97%, SD 11.56;  $t(10) = 4.58, p = .001$ ) and the shape task (Mean 88.95%, SD 5.98;  $t(10) = 21.91, p < .001$ ). This demonstrates that participants were generally responding above chance level on both tasks, even though performance was not matched. There was also no evidence for a correlation between orientation and shape task performance ( $r = .09, p = .788$ ) indicating that different (independent) processing strategies were likely required for each task (i.e. good performance on one task does not necessarily translate to good performance on the other).

### 3.4.2 Hemispheric differences

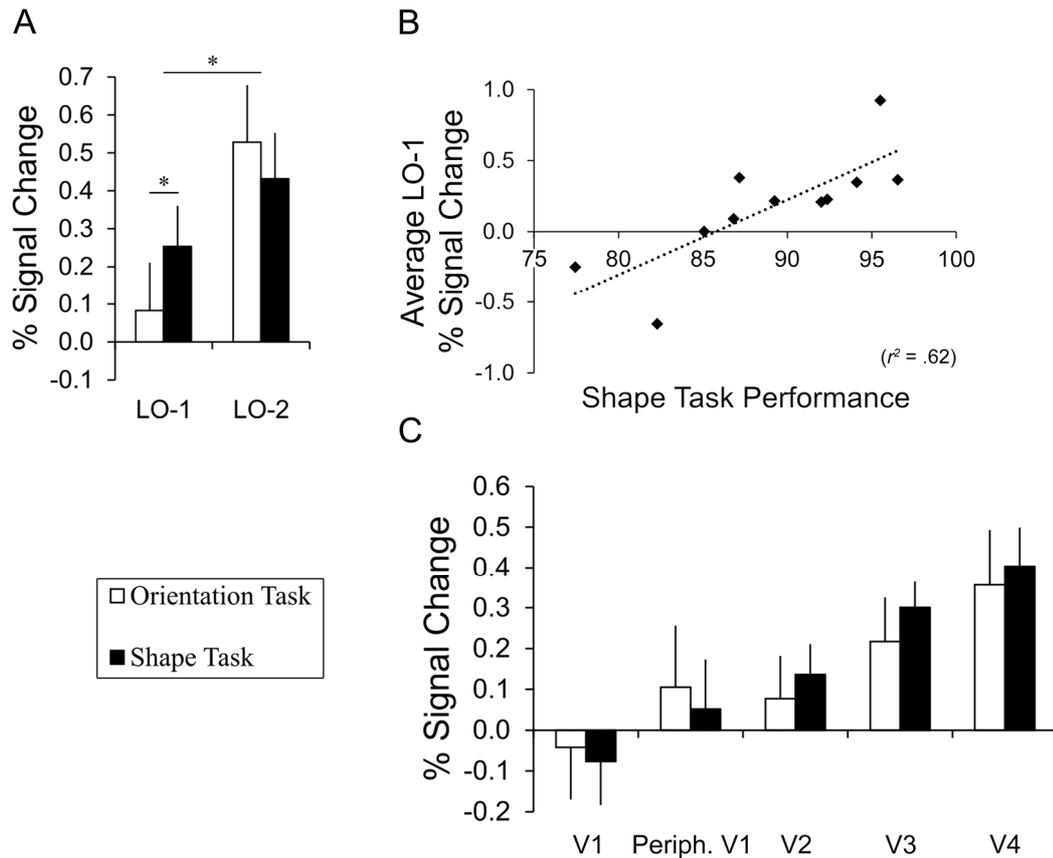
As a preliminary test before our main analysis, we ran a  $2 \times 2 \times 6$  repeated measures ANOVA (task x hemisphere x ROI) to check for either a significant main effect of hemisphere, or any significant interactions involving hemisphere. Mauchly's test indicated that the assumption of sphericity had been violated for the three-way task x

hemisphere x ROI interaction ( $\chi^2(14) = 32.59, p = .004$ ), therefore degrees of freedom were corrected using the Greenhouse-Geisser estimates of sphericity ( $\epsilon = 0.39$ ). This analysis found no significant main effect of hemisphere ( $F(1,10) = 0.25, p = .627$ ), no task x hemisphere interaction ( $F(5,50) = 1.76, p = .214$ ), no hemisphere x ROI interaction ( $F(5,50) = 1.89, p = .114$ ) and no task x hemisphere x ROI interaction ( $F(1.92, 19.29) = 2.04, p = .158$ ). As such, for all subsequent analyses ROIs were collapsed across hemisphere.

### 3.4.3 ROI analysis

To explore results in our key regions of interest, LO-1 and LO-2 (Figure 3.3A), we ran a 2x2 repeated measures ANOVA comparing ROI and condition. A significant main effect of ROI was found ( $F(1,10) = 24.26, p = .001$ ) indicating that activity within LO-2 was generally significantly greater than activity within LO-1. There was no significant main effect of condition ( $F(1,10) = 0.34, p = .571$ ). However, these effects are moderated by a significant interaction ( $F(1,10) = 28.39, p < .001$ ). Paired-samples t-tests indicate that LO-2 was significantly more active than LO-1 for both the orientation task ( $t(10) = 6.93, p < .001$ ) and shape task ( $t(10) = 2.49, p = .032$ ), reinforcing our main effect of ROI. In LO-1, activation was significantly greater for the shape task compared to the orientation task ( $t(10) = -2.30, p = .044$ ), whereas LO-2 showed no significant task differences ( $t(10) = 1.41, p = .189$ ). In sum, the results indicate that LO-2 was generally more active than LO-1 and did not discriminate between tasks, whereas LO-1 did show a significant preference towards the shape processing task.

Intriguingly, LO-1's shape task preference may also be reflected in the behavioural data. Average performance (percent correct) on the shape task correlates highly with percent signal change in LO-1 (Figure 3.3B), not only during the shape task ( $r = .79, p = .004$ ) but also during the orientation task ( $r = .74, p = .010$ ). That is, participants whose LO-1 was *generally* more active tended to show better performance on the shape task. In contrast, shape task performance did not correlate with LO-2's activity during the shape task ( $r = .54, p = .083$ ), and there was only a slight correlation with LO-2's activity during the orientation task ( $r = .65, p = .032$ ). Orientation task performance did not correlate with either ROI under any task (all  $p > .652$ ).



**Figure 3.3.** Summary of Results. (A) Percent signal change in LO-1 and LO-2. LO-1 showed a significant between-task difference but not LO-2, however LO-2 was generally significantly more active than LO-1 for both tasks. (B) Here we averaged activity in LO-1 across both the orientation and shape tasks to highlight the significant correlation between general LO-1 activation and participant's performance on the shape task. (C) Percent signal change in earlier retinotopic ROIs. V1 refers to complete V1, whereas Periph. (peripheral) V1 is restrained to an eccentricity between approximately 7° and 13° visual angle. No between-task differences were observed in any ROI. All error bars represent 1 SEM.

Paired-samples t-tests were also run as a lenient check for between-task differences in lower level visual areas V1-V4 (Figure 3.3C), but no comparisons were close to significance (V1:  $t(10) = 0.42$ ,  $p = .685$ ; V2:  $t(10) = -1.03$ ,  $p = .328$ ; V3:  $t(10) = -1.46$ ,  $p = .174$ ; V4:  $t(10) = -0.66$ ,  $p = .525$ ). It should be noted that there is a trend of increasing activity as you move from V1 towards extrastriate cortex, however this could just be due to larger receptive field sizes yielding greater proportions of active voxels in later regions. As our stimuli were peripheral and retinotopic maps include both foveal and peripheral representations, we did run a final check using our retinotopic ring data to constrain V1 to eccentricities of between approximately 7°

and 13° visual angle. This eccentricity band should better capture our stimuli, as they were 10° into periphery with an average size of 4°. Activation in peripheral V1 was significantly greater than complete V1 for the shape task ( $t(10) = 2.48, p = .033$ ) but it did not quite reach significance for the orientation task ( $t(10) = 1.88, p = .089$ ). However, activity in peripheral V1 still showed no significant differences between tasks ( $t(10) = 0.51, p = .618$ ).

### 3.5 Discussion

To summarise the study and main findings, participants performed orientation and shape comparisons (which shape is most clockwise or spikier respectively) at psychophysically matched thresholds and our intent was to explore activation to identical stimuli under different tasks. LO-2 did not dissociate between the orientation or shape tasks, but did show significantly greater activation for both when compared to LO-1. In contrast, LO-1 appeared to show a significant preference towards the shape task. No between task differences were identified in earlier retinotopic regions (V1-V4), which is unsurprising given that our stimuli were identical across tasks.

First, we must address the performance discrepancy between our two tasks as this potentially undermines any subsequent results. Our original aim was to psychophysically match orientation and shape detection thresholds such that participants would be correct approximately 75% of the time (dependent upon participant's mean responses). However, we then switched to a harder bilateral presentation protocol and consequently decided to use three times participant's thresholds during the scan. We later found that participant's performance was significantly better in the shape task compared to the orientation task. One possibility for this difference is simply that orientation and shape thresholds do not scale in the same way (i.e. their scaling may not be linear), and so tripling thresholds had divergent effects. This remains a possibility, but work outside of this study has run psychophysics using our bilateral protocol, and a similar orientation/shape performance discrepancy was still observed during fMRI. This indicates that our shape task may simply be more susceptible to practice effects than our orientation task (i.e. performance increases due to repeated practice). This could be the case, as orientation judgements must always be *relative*, due to the changing orientation of the reference shape. That is, you need to make comparisons between the two shapes

to complete the task. For the shape task, it is possible that with time participants learn to recognise the profile or characteristics of the spikier and smoother shapes regardless of orientation. In such a scenario a direct comparison may not be necessary and this would explain why the orientation task was adversely affected by bilateral presentation when the shape task was not.

Nevertheless, average performance on both tasks was significantly above chance indicating that in general participants were attending to the stimuli and more importantly, the relevant stimulus properties (i.e. orientation or spikiness). There was also no correlation between performance (percent correct) on the two tasks, indicating that they must require independent strategies. Otherwise, you would expect that participants who performed well on one task would also perform well in the other task. Furthermore, only one region (LO-1) showed evidence of a dissociation between tasks and we have no *a priori* reason to suspect that this region alone was particularly sensitive to the effects of task difficulty. For these reasons, we feel that whilst unfortunate, the discrepancy in task performance does not undermine our main results.

Turning to our findings in LO-1 and LO-2, the results both support and contradict our initial predictions. As hypothesised, LO-2 did show significantly greater activity when compared to LO-1, corroborating findings of it being the more shape-sensitive region (Larsson and Heeger, 2006; Silson et al., 2013). The fact that it did not respond more to the shape task than the orientation task was surprising, given that TMS to LO-2 disrupted a very similar shape-processing task (Silson et al., 2013). As both orientation and amplitude changed across blocks, it is possible that LO-2 was only ever responding to the amplitude differences regardless of task. However, if this were the case we would likely expect greater activation when participants were actively attending to shape amplitude; if anything the opposite was the case. Furthermore, our orientation condition was a very different orientation task to the sinusoidal gratings Silson et al. used, so predictions from one do not necessarily lend themselves to the other.

In a wider sense, LO-2's task insensitivity could support the idea that LO-2 is a preliminary shape processing region for the Lateral Occipital Complex (LOC). There is evidence that the more posterior parts of LOC possess some degree of invariance

to shape manipulations, such as small changes in orientation (Kourtzi et al., 2003). It is likely this also holds for LO-2, given its overlap with posterior LOC (Sayres and Grill-Spector, 2008). This implies that the representation in posterior LOC (and perhaps LO-2) is slightly abstracted from retinotopic space, and so the small manipulations made in this study may not have been registered in that representation. This would explain why no task differences emerged in LO-2, it would also imply that LO-2 is the first region in which a shape representation emerges, given its comparatively robust activation to the shape stimuli regardless of task.

The results for LO-1 were least expected. When Silson et al. (2013) stimulated LO-1 with TMS during a very similar shape processing task to ours, they found that it had no effect, yet here we find that LO-1 shows significant activity for such a task. One explanation for this contradiction is simply that our results demonstrate the inherent flaws in interpreting fMRI data; correlation does not imply causation. The key strength of TMS is it allows tests of causality whereas fMRI can only show a potential link between your experimental manipulation and resulting neural activity. It is possible that LO-1's rising activity may simply be a comorbid response to rising activity in alternative regions, and it may in fact play no crucial role in the task. However, no other region tested showed significant differences across tasks, making this unlikely. We will therefore consider alternative explanations for the discrepancy.

There are two notable differences between our shape processing task and that used by Silson et al. First, we used simultaneous bilateral presentation whereas they used serially presented unilateral stimuli. Second, in terms of orientation, our stimuli were only separated by relatively small fixed amounts, whereas their reference and test stimuli used random orientations. This likely resulted in larger orientation differences than those found in our study, indeed the purpose of their orientation manipulation was to minimise the risks of stimulus overlap, avoiding local luminance cues. It is possible that by having simultaneously presented stimuli at relatively similar orientations, we allowed our participants to identify more local cues to solve the shape processing task. For example, spikier RF patterns have more prominent concavities and steeper, sharper lobes; we know LO-1 plays some role in orientation processing so it could plausibly have detected the local orientations in our RF pattern that cue these differences. In fact, the cross-profile of the D4 (Gaussian) outline used to render our RF contours is not dissimilar to part of a sinusoidal

grating. As the orientation task still needs relative judgements across both shapes, it may have encouraged more global processing; just comparing the ‘gist appearance’ of the two shapes. Naturally this is speculative, it is impossible to determine the individual criteria our participant’s used to solve each task. However, this explanation is in accordance with the earlier explanation for practice effects in the shape task but not the orientation task. If local orientation cues informing the steepness of an RF pattern’s lobe could be detected, participants could potentially learn to use this without requiring a relative comparison. Furthermore, if true then participants with better performance on the shape task would probably be making more use of local cues, explaining the correlations between shape task performance and LO-1 activity. Finally, the idea also fits with the possibility that differing receptive field sizes may have driven Silson et al.’s dissociation. LO-1’s smaller receptive fields (Amano et al., 2009) would likely make it more tuned for the local orientation differences in a lobe over more global orientation comparisons.

To conclude, this study used a block design experiment with two conditions; orientation and shape. The stimuli were identical between conditions, the only difference was that in the former condition participants performed an orientation judgement task, whereas in the latter participants performed a shape (or curvature) judgement task. In this manner, we aimed to test whether orientation selectivity in LO-1 remains regardless of the medium used to carry that orientation information. We found that in line with previous reports, LO-2 was generally the more shape selective region. This is the ROI that also shows greatest overlap with object-selective LOC (Sayres and Grill-Spector, 2008; Silson et al., 2016), so shape sensitivity here is expected. In contrast to our predictions, we actually found LO-1 showed significant preference towards our shape processing task, over the orientation task. We propose that this finding reflects differences in the strategies used to solve our two tasks; whereas orientation comparisons require relative judgements likely based upon the global appearance (or ‘gist’) of a shape, shape comparisons could feasibly be solvable from local information alone. As such, from the evidence so far it seems plausible that LO-1’s role is to make more local, internal judgments about shape (or other visual stimuli), whereas LO-2 pools information over a wider area to process global shape information.

## **Chapter 4. Responses to Shape from Form Coherence in the Lateral Occipital Cortex**

### **4.1 Abstract**

Previous research has found greater sensitivity to second-order stimuli in retinotopic visual area LO-1, compared to LO-2. However, these studies used edge or grating stimuli and LO-2 is more attuned to shape information. As such, we hypothesised that shapes defined by second-order cues may be more likely to elicit activity in LO-2. We assessed responses to fixed- and random-orientation shapes defined by form coherence, compared to a 'baseline' condition with no regions of coherence. As expected, LO-2 showed greater activation than LO-1. It also responded more strongly to the random- than fixed-orientation shapes implying that its responses were specifically due to second-order shape sensitivity, as the degree of form coherence was matched across conditions. Similar results were observed in functionally-defined LO and pFs, but not lower-level retinotopic regions V1-V4. In sum, we corroborate findings of shape sensitivity in LO-2, whilst also extending previous work to show that LO-2 will respond to second-order stimuli.

## 4.2 Introduction

In this chapter, we aim to characterise how retinotopic visual areas LO-1 and LO-2 (Larsson and Heeger, 2006; Wandell et al., 2007) respond to second-order contours, that is, contours not defined through luminance modulation.

There is evidence for second-order sensitivity in LO-1. Larsson, Landy and Heeger (2006) used gratings defined with standard luminance modulation (i.e. black and white bars), plus second-order versions with modulations in contrast (i.e. high contrast and low contrast bars) and orientation (i.e. bars of broadly vertical orientations intersected with bars of broadly horizontal orientations). After adaptation, LO-1 showed significantly more activity to orthogonal than parallel gratings, regardless of how those gratings were defined. This indicates a release from adaptation, and therefore second-order sensitivity. Similarly, Montaser-Kouhsari, Landy, Heeger, & Larsson (2007) found a release from adaptation in LO-1 for orthogonal illusory edges (defined using misaligned gratings).

Notably, these papers reported only limited second-order sensitivity in LO-2 relative to LO-1, observations that could in part arise due to the choice of either edge or grating stimuli. Larsson and Heeger originally noted that whilst LO-1 showed sensitivity to grating orientation, LO-2 responded preferentially to shapes. Furthermore, Silson et al. (2013) found an LO-1/LO-2 double dissociation; transcranial magnet stimulation (TMS) to LO-1 impaired grating orientation judgements, whereas TMS to LO-2 impaired shape processing judgements. Prompted by this, we questioned whether LO-2 would show second-order sensitivity to shape rather than edges or grating stimuli.

The response of LO-2 to second-order stimuli is of key importance in the context of shape processing as LOC (Malach et al., 1995; Grill-Spector et al., 2001) is known to have second-order sensitivity. For example, LOC responds to shapes defined using texture discrepancies (Grill-Spector et al., 1998a) or illusory contours based upon Kanizsa-style stimuli (Murray et al., 2002a; Ritzl et al., 2003; Stanley and Rubin, 2003). Therefore, if LO-2 fails to respond to second-order shape stimuli while LOC does, then at the very least it implies alternative processing pathways might exist dependent upon where second-order shape sensitivity arises.

To investigate, second-order shape stimuli were created using form coherence. Specifically, we contrasted a shape defined with coherently orientated lines against a background of randomly orientated lines. Stimuli were presented using a block-design fMRI experiment, with the shapes presented at either fixed or random orientations within a block. They were contrasted against a 'baseline' condition with no regions of coherence.

This design was in part based upon adaptation; there is evidence that rotation can release LOC from adaptation (Grill-Spector et al., 1999) and so on this basis we would expect greatest activation for our random-orientation condition. However, this is naturally predicated upon detection of the shape. The degree of form coherence is equivalent across fixed- and random-orientation shape conditions, and so those two conditions should be equivalent for regions without second-order shape sensitivity. As such, greater activation during the random-orientation shape condition compared to the fixed-orientation shape condition should imply second-order shape sensitivity.

We expect that LOC will show second-order shape sensitivity, in-line with the aforementioned literature. Critically, based predominantly upon Silson et al. (2013) we also predict that LO-2 should show greater second-order sensitivity than LO-1 due to the use of shape stimuli. Such a finding would contrast previous evidence of greater second-order sensitivity in LO-1 (Larsson et al., 2006; Montaser-Kouhsari et al., 2007) and also reinforce the orientation/shape dissociation between these two regions.

### **4.3 Methods**

#### **4.3.1 Participants**

Seventeen participants (ages 20-47, all gave informed consent) were recruited from the University of York Psychology department. Two participants were later excluded from data analysis due to excessive motion during the functional scan, one participant was excluded due to poor eyesight (due to a large cylindrical correction, the visual acuity was <6/6 when corrected in the scanner with spheres alone).

In addition to a functional scan, high-resolution structural images and retinotopic-mapping data were acquired for all participants. For six participants, we had previously acquired functional localiser data to identify the LOC. These structural,

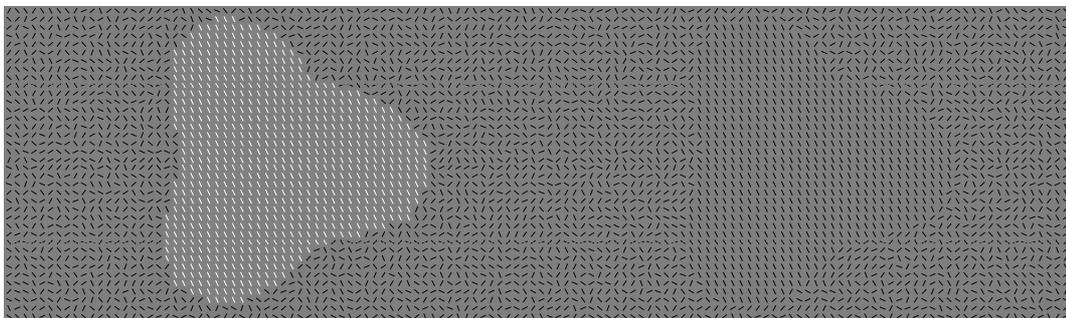
retinotopic and localiser scans were carried out and analysed as described in the Methods chapter (Chapter 2, sections 2.3 to 2.6).

### 4.3.2 Stimulus Presentation

All stimuli were rear projected onto an acrylic screen in the bore of the scanner, at a resolution of 1024x768 (40x30 degrees of visual angle). Lying supine, participants viewed the stimulus via a front-silvered mirror placed over the head coil. Matlab (Mathworks Matlab R2012a) with the Psychophysics toolbox v3.0.10 (Brainard, 1997; Pelli, 1997) was used for stimulus generation and presentation.

### 4.3.3 Functional MRI data

To create our second-order stimuli, the screen was segmented into a grid of squares ( $0.5^\circ \times 0.5^\circ$  visual angle per square), with a black line ( $0.24^\circ \times 0.04^\circ$  visual angle) centred in each. These lines were pseudo-randomly orientated, having angles of between  $20^\circ$ - $70^\circ$  or  $110^\circ$ - $160^\circ$ . Cardinal orientations (horizontal or vertical) were avoided due to the oblique effect (Appelle, 1972); a shape consisting of purely horizontal or vertical lines could have been atypically salient, leading to activation above and beyond that expected purely from form coherence alone. The lines were not spatially jittered as preliminary testing indicated that this would make any shape stimuli almost impossible to detect. The background of the stimuli was set to mid-grey (RGB 128,128,128). Due to the limited number of pixels available to each line, the stimulus was rendered at eight times the desired resolution, then down-sampled and anti-aliased using Lanczos-3 kernel interpolation. This ensured that all lines had a smooth and consistent appearance.



**Figure 4.1.** Example second-order shape stimulus. The shape (three-lobed radial frequency pattern) on the left is highlighted for illustrative purposes, the shape on the right is depicted as it would appear to participants. Note that the randomly orientated lines in the background extend to the full extent of the screen during actual stimulus presentation.

In the baseline condition, the entire screen consisted of lines with no regions of coherence. For the fixed- and random-orientation conditions, bilateral regions of coherence were used to define shapes as shown in Figure 4.1. The shape chosen was a three-lobed radial frequency (RF) pattern, generated through sinusoidal modulation of a circle's perimeter with the following formula:

$$4.8 \times (1 + 0.25 \sin(3\theta + \phi))$$

Here, 4.8 specifies the circle's radius (or average RF pattern radius) in degrees of visual angle,  $\phi$  gives the orientation of the resultant RF pattern and was pseudo-randomly assigned as above (i.e. avoiding orientations that were vertically or horizontally symmetric).

An RF pattern was chosen for two reasons. Firstly, TMS to LO-2 disrupted the processing of an RF pattern's shape profile (Silson et al., 2013), suggesting LO-2 is sensitive to this type of stimulus. Secondly, by their very nature RF patterns have a broadly circular outline. This means that rotation will have minimal effect on the spatial location of coherence.

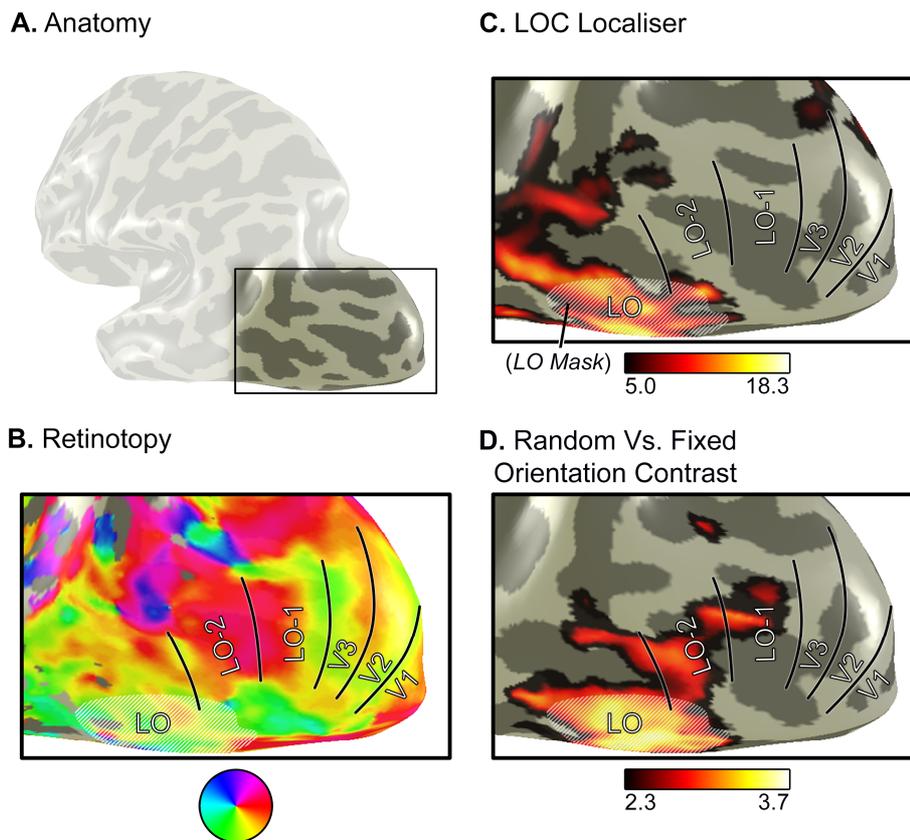
Two such RF patterns were centred vertically and placed along the horizontal meridian either side of fixation (the centre of each pattern was  $10^\circ$  from fixation). All lines whose centre fell within this pattern were allocated coherent orientations, all other background lines were oriented pseudo-randomly as in baseline. Note that the mask was not mirrored about fixation, as this would have created symmetry cues not present in baseline.

All conditions were presented in 10sec blocks, and the orientation of *every* line changed once per second. In the fixed-orientation condition, the orientation of the RF shape was fixed for a given block, but changed between blocks. In the random-orientation condition, the RF shape's orientation changed once per second. No stimulus-related task was included, but participants were asked to maintain fixation on a central red cross. Blocks were presented in a predetermined, cyclical order, such that a given condition was equally likely to occur after any other condition, but blocks of the same condition were not repeated sequentially. Sixteen blocks of each condition were included, leading to 480sec of scanning time. No blank periods were included between or within blocks to avoid sudden changes in luminance.

## 4.4 Data Analysis

### 4.4.1 Retinotopy

The retinotopic maps were defined using standard procedures (see ‘Chapter 2: Methods’ section 2.6 for details). Briefly, by visualising retinotopic data on the inflated surface of each participant (Figure 4.2A) we could identify reversals or discontinuities in the retinotopic map. This allows us to demarcate visual regions of interest (ROIs) V1-V4, LO-1 and LO-2 in all participants (Figure 4.2B).



**Figure 4.2.** Regions of interest and functional activity for one representative participant. (A) Left Occipital area of interest (highlighted), subsequently magnified in panels B-D. (B) Visual field representation of the cortex (see colour key below panel), with visual area boundaries (horizontal and vertical meridians) shown in black. (C) Activity elicited by an objects versus scrambled objects localiser, Z-Score thresholded at  $Z > 5.0$ . The largest contiguous cluster of activity was extracted as the Lateral Occipital Complex (LOC), then manually split into dorsal/posterior LO and ventral/anterior pFs (*not shown*). Also depicted, the LO mask (white shaded region) defined in group space (see Methods, section 4.4.3), back-transformed into individual space. (D) Group-level functional activity elicited by the random- versus fixed-orientation contrast, Z-Score thresholded at  $Z > 2.3$  and again back-transformed into individual space, for illustrative purposes.

#### 4.4.2 Functional Data

Functional data were analysed using FEAT (fMRI Expert Analysis Tool) (Worsley, 2001). At the individual level (fixed effects analysis), a high-pass filter cut-off point of 50sec was used, to remove low-frequency drift. Motion was corrected for (using MCFLIRT), and the estimated motion parameters were entered as confound covariates. For two participants, additional covariates were generated using FSL's motion outliers tool to model out the effects of motion spikes. Spatial smoothing was set to 6mm (twice voxel size) and FILM prewhitening was used. Data were first aligned to each participant's average 16-channel T1, and then to standard space (MNI 152 T1 2mm brain; linear alignment). Contrasts were created for both fixed- and random-orientations over baseline, random-orientations over fixed-orientations and vice versa.

At the group level (mixed effects analysis; FLAME 1), the Z-Stats were cluster-corrected using a Z threshold of 2.3 and a cluster  $p$  threshold of .05. The results of this analysis are shown in Figure 4.2D, but these will be discussed in more detail later in the Results section.

All ROI analysis was performed at the individual level, using FEATquery on the COPE (contrast of parameter estimates) statistics for fixed- and random-orientation conditions over baseline. COPE values were converted into percent signal change. Initial statistical tests (two-tailed paired-samples t-tests) were run to check for hemispheric differences across all ROIs. Only V1 showed marginal differences (fixed-orientation:  $p = .012$ ; random-orientation:  $p = .058$ ), all other comparisons were non-significant (all  $p > .163$ ). As this result would not survive correction for multiple comparisons and we had no *a priori* reason to believe hemispheric differences should exist, ROIs were collapsed across hemisphere for all subsequent analyses.

#### 4.4.3 Localiser

Analysis of the LOC functional localiser at the individual level was similar to the functional data procedure. As before, spatial smoothing was set to twice voxel size (in this case 4mm), the high pass filter was set to twice the length of one object block, plus one scrambled object block (60sec). A contrast was set up to compare objects over scrambled objects. At the second level (combining each participant's

three runs) fixed effects analysis was used and the Z-Stats were again cluster-corrected, although a more stringent threshold was used ( $Z > 5.0$ ,  $p = 1 \times 10^{-6}$ ) due to the large amount of data collected per participant. The largest contiguous cluster in each hemisphere that responded significantly more to objects than scrambled objects (Figure 4.2C) was then extracted and manually divided into LOC's more posterior/dorsal component LO, and its more anterior/ventral component pFs (posterior fusiform gyrus).

As an additional measure, we ran a third level analysis (mixed effects analysis; FLAME 1) on the localiser data to combine results across participants and transform it to standard space. This used cluster-correction with a Z threshold of 2.3 and a cluster  $p$  threshold of .05. The main clusters identified at this stage were again manually divided into bilateral LO and pFs ROIs (Figure 4.2C). To distinguish these group ROIs from ROIs specific to a given individual, we will describe these as the LO mask and pFs mask respectively. The reason for this analysis was to generate estimated LO and pFs ROIs for all individuals, while having functional data for only six of the participants.

## 4.5 Results

### 4.5.1 Responses in early retinotopic visual areas V1-V4

We first considered how early visual areas V1-V4 responded, as effects here could potentially influence interpretation of our key ROIs LO-1 and LO-2. The differences in the cortical responses of V1-V4 to our fixed- and random-orientation shape stimuli compared to baseline are plotted in Figure 4.3A. There appears to be a progressive increase in response to the stimuli from primary visual cortex to V4. Interestingly, in V1 and to a lesser extent V2 the signals elicited by the shape stimuli are smaller than responses to the baseline condition in which no coherence was present. It is only in V4 where signals elicited by the shape exceed those elicited by the baseline stimuli.

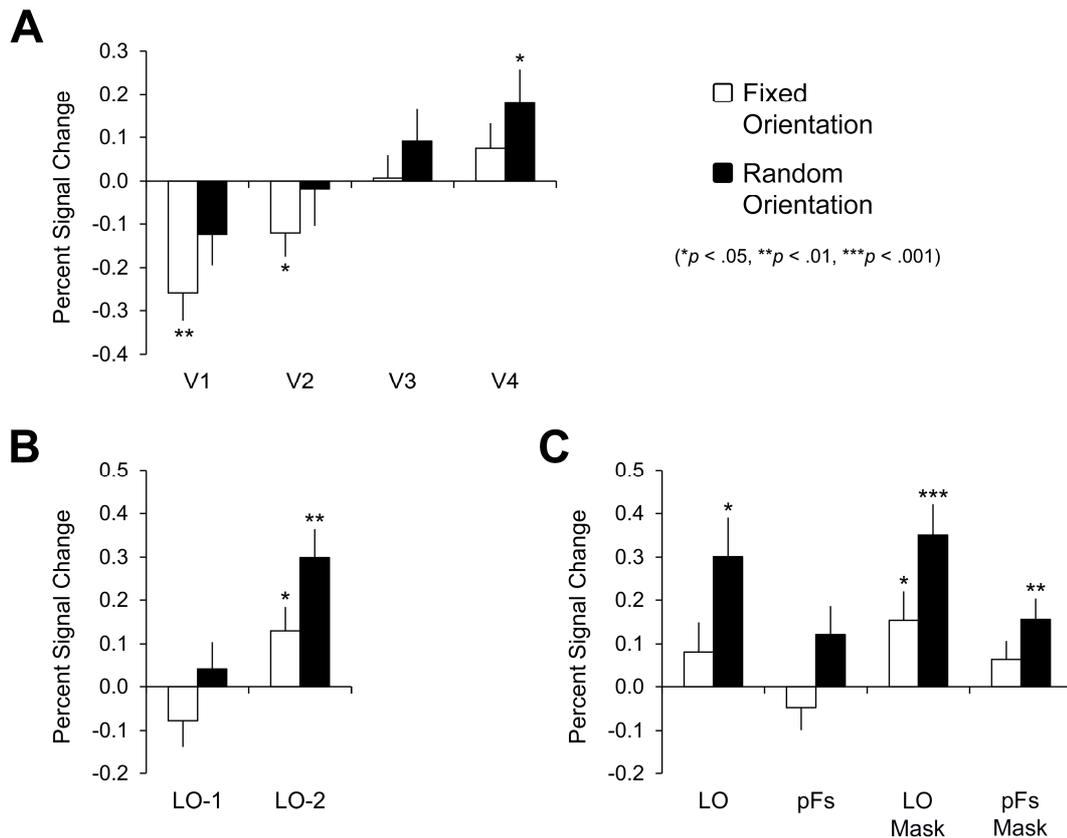
To check for significant differences or interactions between visual areas and stimulus conditions, we ran a 2x4 repeated-measures ANOVA (Stimulus Condition x ROI). Mauchly's test indicated that the assumption of sphericity had been violated for the main effect of ROI ( $\chi^2(5) = 13.75$ ,  $p = .018$ ) and Condition x ROI interaction ( $\chi^2(5) = 15.22$ ,  $p = .010$ ), therefore degrees of freedom were corrected using the

Greenhouse-Geisser estimates of sphericity ( $\epsilon = 0.58$ ;  $\epsilon = 0.55$  respectively). Whilst from Figure 4.3A there appears to be a trend of greater activity for the random- over fixed-orientation conditions, the main effect of condition did not reach significance ( $F(1, 13) = 3.96, p = .068$ ). A significant main effect of ROI was found ( $F(1.74, 22.68) = 33.51, p = 3.7 \times 10^{-7}$ ), there was no significant interaction ( $F(1.65, 21.39) = 0.93, p = .394$ ).

To explore the main effect of ROI we ran Bonferroni-corrected post-hoc tests, which showed all between-ROI comparisons were significant (all  $p < .035$ ). This reflects the general trend of increasing activity when moving from V1 into extrastriate cortex. Of key importance here is the lack of significant main effect of condition, as this implies that any dissociations identified further in the visual stream are unlikely to be due to lower-level stimulus differences.

#### 4.5.2 Responses in Lateral Occipital retinotopic visual areas LO-1 and LO-2

In Figure 4.3B we show the signals elicited by the shape stimuli in our main regions of interest LO-1 and LO-2. LO-2 appears to generally respond more than LO-1, and there also appears to be a stronger response to the random- compared to fixed-orientation shapes, most notably in LO-2. To test for differences in these signals we applied a 2x2 repeated-measures ANOVA (Stimulus Condition x ROI). A significant main effect of ROI was found ( $F(1,13) = 30.67, p = 9.6 \times 10^{-5}$ ), confirming the greater response in LO-2 when compared to LO-1. There was also a significant main effect of condition ( $F(1,13) = 13.10, p = .003$ ), suggesting both regions responded significantly more to the random- than fixed-orientation shapes. There was no significant ROI-condition interaction ( $F(1,13) = 2.98, p = .108$ ). However, it should be noted that LO-1's activity was effectively at baseline as can be seen in Figure 4.3B, only LO-2 demonstrated robust activity to the shape stimuli. Supporting this, two-tailed one-sample t-tests find no significant differences from baseline in LO-1 for the fixed- ( $t(13) = -1.29, p = .221$ ) or random-orientation ( $t(13) = 0.63, p = .539$ ) conditions, whereas LO-2 differs significantly from baseline in both (fixed-orientation:  $t(13) = 2.39, p = .033$ ; random-orientation:  $t(13) = 4.58, p = .001$ ).



**Figure 4.3.** Percent signal change for different brain regions. Note that all activity is plotted relative to ‘baseline’; activity elicited by an array of incoherently orientated lines. (A) Early retinotopic regions V1-V4 show a rising trend of activity but no general main effect of condition. (B) Retinotopic regions of primary interest, LO-1 and LO-2. Whilst activity within LO-1 is effectively at baseline, LO-2 shows robust activation for the fixed- and (to a greater extent) random-orientation conditions. (C) Activity in functionally-defined LO and pFs (six participants only), as well as activity in LO and pFs masks (for all participants) created in group space. Again, robust activity is observed, primarily in the random-orientation condition. Error bars in all graphs represent one standard error of the mean. Significance values are derived from (uncorrected) two-tailed one sample t-tests to highlight activity that differs significantly from zero. Full statistical analysis for the data in each panel is presented in the main text.

### 4.5.3 Responses in functionally-defined visual areas LO and pFs

Having seen a selectivity to shape defined by second order cues in LO-2, we sought to also confirm this selectivity in functionally defined extrastriate areas, LO and pFs. In Figure 4.3C we show the responses to our shape stimuli in LO and pFs, first, in those six participants for whom we have localised these areas, and second, in all participants based upon LO and pFs masks. There is a striking similarity between the

responses found in LO and LO-2, while responses overall in pFs seem lower than in LO.

A 2x2 repeated-measures ANOVA (Stimulus Condition x ROI) was run to compare activity in LO and pFs for those six participants with localiser data (Figure 4.3C). The main effects of condition was close to significance ( $F(1,5) = 6.38, p = .053$ ), no main effects of ROI ( $F(1,5) = 4.86, p = .079$ ) or Condition x ROI interactions ( $F(1,5) = 2.61, p = .167$ ) were observed. However, this statistical test was underpowered due to the limited number of participants in this part of the study.

In an attempt to address this, we looked for significant activity in the LO and pFs masks created in group space (Figure 4.3C). First, to check their validity we ran four two-tailed paired-samples t-tests as a means to compare activity in the original LO and pFs ROIs to those created in group space (in only the participants with localiser data). No significant differences were found, indicating that our mask ROIs are reasonable approximations of their original counterparts (fixed orientation, LO/LO mask:  $t(5) = 1.03, p = .353$ ; pFs/pFs mask:  $t(5) = 0.67, p = .533$ ; random orientation, LO/LO mask:  $t(5) = 1.01, p = .358$ ; pFs/pFs mask:  $t(5) = 1.45, p = .207$ ). We then ran a 2x2 repeated-measures ANOVA on these mask ROIs as above. The main effect of condition now reaches significance ( $F(1,13) = 19.06, p = .001$ ) suggesting generally greater activity to the random- over fixed-orientation shapes. A significant main effect of ROI ( $F(1,13) = 10.04, p = .007$ ) and a significant Condition x ROI interaction ( $F(1,13) = 11.13, p = .005$ ) also emerged. This implies that the region within the LO mask was generally more active and also showed greater between-condition differences when compared to activity within the pFs mask.

#### 4.5.4 Whole-brain analysis

Finally, we explored the results of a whole-brain analysis. The fixed-orientation over baseline contrast identified one significant cluster in the right hemisphere (MNI centre of mass co-ordinates (mm): 49.50, -72.70, 3.49) whereas bilateral clusters were identified in both the random-orientation over baseline contrast (Left: -31.60, -78.10, 5.47; Right: 37.80, -71.20, 3.90) and the random- versus fixed-orientation contrast (Left: -39.50, -78.90, -0.45; Right: 37.30, -74.30, 6.02; Figure 4.2D).

The latter two clusters showed considerable overlap with the LO mask ROIs defined in group space; 96.7% of the voxels in the left LO mask (-42.95, -78.06, -6.17) and 80.7% of the voxels in the right LO mask (43.31, -74.00, -9.52) lie within the clusters. This strongly indicates that the primary region identified in the random-over fixed-orientation contrast was LO. We also compared overlap with our LO-1 and LO-2 maps in standard space. On average, 35.5% (SD 16.9%) of voxels within LO-1 and 70.2% (SD 14.0%) of voxels within LO-2 were within the clusters. To compare, just 4.4% of voxels within LO-1 and 15.5% of voxels within LO-2 lay within our group LO mask on average. This implies that sensitivity to our second-order shapes is likely emerging within LO-1 and LO-2, before continuing to LO. Indeed, this is also suggested when we overlay the random- versus fixed-orientation activation on the anatomy of our representative participant (Figure 4.2D).

## 4.6 Discussion

To summarise the main findings; second-order shape selectivity emerged in LO-2, with both fixed- and random-orientation conditions eliciting significant levels of activity. Similar levels of shape selectivity were present in LO and (to a lesser extent) pFs. No significant differences between conditions were found in earlier retinotopic areas (V1-V4), however a general trend of rising activity when moving from V1 towards extrastriate cortex was observed.

Our results support the findings of Montaser-Kouhsari et al. (2007), who showed that both LO-1 and LO-2 have some sensitivity for illusory contours, but they also extend them. Past research suggested that LO-2 may be less sensitive to second-order or illusory boundaries when compared with LO-1, but this work focussed primarily on straight edges and grating stimuli (Larsson et al., 2006; Montaser-Kouhsari et al., 2007). Here we found the opposite result; LO-2 was significantly more responsive to our second-order stimuli when compared with LO-1.

Furthermore, LO-2's responses to the random-orientation condition were significantly greater than its responses to the fixed-orientation condition, implying that the activity was specifically cued by shape information. This supports our prediction that shape information is necessary to elicit activity in LO-2, it also corroborates Silson et al.'s (2013) finding of LO-2, not LO-1, being the shape sensitive region. In LO-1, whilst a significant main effect of condition emerged, its

responses were nevertheless effectively at baseline. This implies that second-order cues alone are not sufficient to elicit robust activity in this region.

Our secondary hypothesis, that LO would respond to shapes defined from form coherence, was also supported. When considering the LO and pFs masks in group space, a strong significant main effect of condition was identified suggesting that these regions are sensitive to our shape stimuli (particularly so for LO). A similar trend of activity was identified in participants with individual localiser data, corroborating the finding. Furthermore, the LO masks showed substantial overlap with the clusters of activation from a whole brain random- over fixed-orientation contrast. Together, this evidence strongly implies that LO is the region primarily activated by our stimuli. Generally this is a relatively unsurprising result, LO is a known shape-responsive area (Malach et al., 1995; Grill-Spector et al., 2001) and it has previously been shown to respond to illusory contours (Murray et al., 2002a; Ritzl et al., 2003; Stanley and Rubin, 2003). This finding serves more to legitimise the stimuli and consequently the results for LO-2.

Results in earlier retinotopic areas V1-V4 also serve to support the idea that second-order shape sensitivity only truly emerges in LO-2. Whilst there is a trend towards greater activity for the random- over fixed-orientation stimuli in the early areas, it is far less prominent than that found in LO-2 onwards. This is demonstrated by the lack of significant main effect of condition in V1-V4, indicating that variance across conditions was large enough to offset any general trends. Notably, the pattern of rising activity when moving from V1 towards extrastriate cortex is the opposite of that found when using edge or grating stimuli (Larsson et al., 2006; Montaser-Kouhsari et al., 2007). This could imply that different neural mechanisms underlie responses to second-order lines and the second-order shapes used here.

Finally, V1 showed an unexpected preference for incoherently orientated lines compared to our main conditions containing regions of coherence (i.e. responses to our key conditions were below 'baseline'). We consider two explanations for this result. First, the regions of coherence may have led to surround suppression (e.g. Jones et al., 2001). A given line within the coherent region would be surrounded by a multitude of identical lines outside its classical receptive field. This could have produced a suppressing effect, and therefore elicited the diminished activity when

compared to arrays with no coherent regions. Alternatively, top-down effects could play a role. There is evidence that perceptually grouping elements into a coherent shape causes increased activity in LO, and decreasing activity in V1 (Murray et al., 2002b; Fang et al., 2008). Therefore, when our coherent lines were grouped to form a coherent perception of a shape this may have inhibited V1. However, we see greater activity in LO for the random- over fixed-orientation condition, but this is not accompanied by a greater reduction of activity in V1. As such, this is not a perfect explanation but it remains a possibility.

In summary, our findings provide converging evidence for greater shape-sensitivity in LO-2, when compared to LO-1. We also extend prior research, as we show that LO-1 is not simply more sensitive to second-order cues than LO-2, the choice of stimulus is critical. Here, using second-order shapes, we find that it is actually LO-2 with greater second-order sensitivity.

## **Chapter 5. Multivariate Patterns in the Human Object-Processing Pathway Reveal a Shift from Retinotopic to Shape Curvature Representations in LO-1 and LO-2**

### **5.1 Abstract**

Representations in early visual areas are organised on the basis of retinotopy, however this organisational principle appears to lose prominence in extrastriate cortex. Nevertheless, an extrastriate region such as shape-selective Lateral Occipital Cortex (LO) must still base its activation upon the responses from earlier retinotopic visual areas, implying a transition from retinotopic to ‘functional’ organisations should exist. We hypothesised that such a transition may lie in LO-1 or LO-2; two visual areas lying between retinotopically-defined V3d and functionally-defined LO. Using a rapid event-related fMRI paradigm, we measured *neural* similarity in 12 human participants between pairs of stimuli differing along dimensions of shape exemplar and shape complexity within both retinotopically- and functionally-defined visual areas. These neural similarity measures were then compared to low-level and more abstract (curvature-based) measures of *stimulus* similarity. We found low-level, but not abstract stimulus measures predicted V1-V3 responses, while the converse was true for LO – a double dissociation. Critically, abstract stimulus measures were most predictive of responses within LO-2, akin to LO, whilst both low-level and abstract measures were predictive for responses within LO-1 perhaps indicating a transitional point between those two organisational principles. Similar transitions to abstract representations were not observed in the more ventral stream passing through V4 and VO-1/2. The transition we observed in LO-1 and LO-2 demonstrates that a more ‘abstracted’ representation, typically considered the preserve of ‘category-selective’ extrastriate cortex, can nevertheless emerge in retinotopic regions.

## 5.2 Introduction

As discussed in the introductory chapter, there are two prominent means to identify visual areas of the human brain. They can be identified by their retinotopic representations of the visual field (Engel et al., 1994; Wandell et al., 2007) or by their selectivity to stimulus categories (Malach et al., 1995; Kanwisher et al., 1997; Epstein and Kanwisher, 1998; Downing et al., 2001). These differing approaches persist largely because category-selective areas appear to have far weaker retinotopy than early visual areas (Sayres and Grill-Spector, 2008), implying such areas might map other (more ‘abstract’) dimensions of stimulus content (Op de Beeck et al., 2008a).

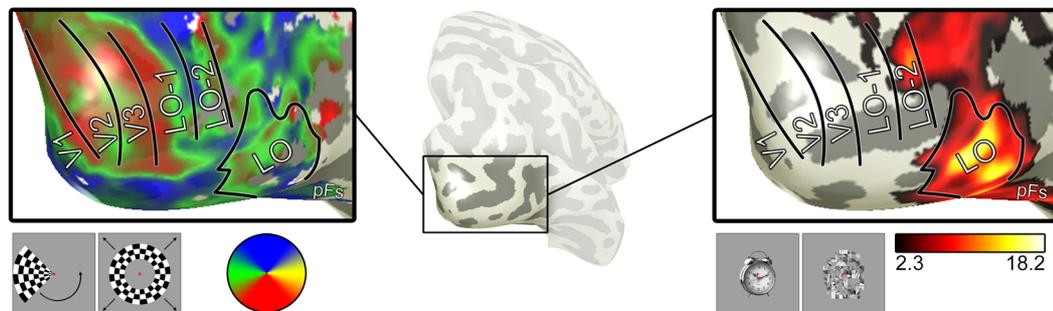
There is considerable interest in the relationship between early, retinotopic representations and representations in later extrastriate cortex (Levy et al., 2001; Hasson et al., 2002; Malach et al., 2002; Larsson and Heeger, 2006; Hemond et al., 2007; Sayres and Grill-Spector, 2008; Schwarzlose et al., 2008; Arcaro et al., 2009; Grill-Spector and Weiner, 2014). In this chapter, and those that follow, we considered this relationship in greater detail. Specifically, we explored how topographic representations of the visual field shift to representations of more abstract stimulus properties, and whether these organisational principles are mutually exclusive. Note that ‘abstract’ is used throughout this chapter to encompass representations that are abstracted from retinotopic representations predominating in earlier visual areas. We specifically investigated ‘curvature complexity’ as a potential abstract feature that extrastriate areas may be tuned to, asking where in the visual hierarchy such representations might emerge.

It is well-established that the Lateral Occipital Complex (LOC) is selective for shapes over stimuli without coherent form (Malach et al., 1995; Grill-Spector et al., 2001). This region has been divided into more posterior/dorsal Lateral Occipital Cortex (LO) and more anterior/ventral Posterior Fusiform Sulcus (pFs), however their underlying organisation remains a source of debate (Op de Beeck et al., 2008a). Multiple studies find activity patterns within LO correspond to similarity between shape features. For example, Op de Beeck et al. (2008b) found perceptual shape similarity (cued predominantly by shape features, e.g. protrusion spikiness) corresponded with LO activity. Similarly, patterns of activity in LO correlate with measures of aspect ratio and skew (Haushofer et al., 2008b), and the prominence of

protrusions in radial frequency patterns (Drucker and Aguirre, 2009). Generally, these studies found strictly physical similarity measures (e.g. pixel-wise similarity) were poor predictors of LO activity, implying a move from retinotopic organisation towards something more abstract.

This transition is unlikely to be abrupt. Retinotopy is the currency of early visual cortex, therefore retinotopy must ultimately be the foundation for higher-level representations (Peirce, 2015). This implies the existence of an intermediate stage where retinotopic and more abstract representations coexist. We asked where this point might lie.

Two candidate regions for such a point are the retinotopically-defined visual field maps LO-1 and LO-2 (Larsson and Heeger, 2006), lying nearby or overlapping with LO (Sayres and Grill-Spector, 2008; Silson et al., 2016). LO-2 in particular responds well to objects (Figure 5.1), and is causally involved in shape-processing tasks (Silson et al., 2013), implying some similarity with LO. This lateral occipital aspect of the human brain therefore offers an ideal site to explore how organisational principles transition from retinotopic to abstract.



**Figure 5.1.** Identification of visual areas. Centre: the occipital area of interest (highlighted) from one representative participant, that is magnified in the panels to the Left and Right. In the Left panel the visual field representation of the cortex is shown (see colour key below). Boundaries between visual areas are shown in black and correspond to the representations of the horizontal and vertical meridians. Data are given for responses to a rotating wedge (as inset below the data). Also shown is the outline of LO and general location of pFs, which is derived from an objects versus scrambled objects functional localiser. Data here are Z-score thresholded at  $Z = 2.3$ , however LO itself was defined using a sphere surrounding peak voxel (see Materials and Methods).

We hypothesised that representations in V1-V3 would be driven strongly by low-level (retinotopic) similarity, whereas LO should be driven by more abstract measures of shape curvature. For LO-1 and LO-2 (lying between V3d and LO), their full hemifield representations imply retinotopic tunings, yet their overlap with LO implies more abstracted representations may also exist. Therefore, we hypothesised a retinotopic-to-abstract transition may occur near or in LO-1 and LO-2.

To test our hypotheses, we created stimulus sets differing along dimensions of shape exemplar (e.g. bird or cat), and level of shape detail (i.e. how complex the shape outline is). BOLD responses to each of our stimuli were recorded under a rapid event-related fMRI design, and measures of neural similarity were extracted for the visual areas discussed above. These *neural* similarity measures could then be compared to various *stimulus* similarity measures; to reveal which stimulus features are most salient for each visual area.

### **5.3 Materials and Methods**

#### **5.3.1 Participants**

Twelve participants (Mean age 25.42, SD 4.78; 8 Males; all gave informed consent) were recruited from the University of York Psychology department. Each participant underwent one high-resolution structural scanning session, one retinotopic-mapping session, one localiser session and two main functional sessions, totalling 5.25 hours scanning per participant. In addition, all participants performed two 30 minute behavioural sessions.

#### **5.3.2 Preliminary Data Acquisition and Analysis**

All imaging data were acquired using a GE 3-Tesla Sigma HD Excite scanner and a 16-channel head coil, to improve signal-to-noise in the occipital lobe. All participants underwent standard structural and retinotopic mapping scans as described in Chapter 2 (Methods, sections 2.3 and 2.6). Retinotopy data were analysed using standard techniques (Wandell et al., 2007) as specified previously (Baseler et al., 2011). Regions of interest (ROIs) V1-V4, LO-1 and LO-2 (Larsson and Heeger, 2006) were identified in both hemispheres of all participants.

Participants also underwent LOC localiser scans, contrasting objects versus scrambled objects (again as described in Chapter 2, Methods, section 2.5). The

localiser data were analysed using FEAT (FMRI Expert Analysis Tool; Worsley, 2001). At the first (individual) level we removed the first three volumes and used a high-pass filter cut-off point of 60sec to correct for low-frequency drift. Spatial smoothing was performed with a Gaussian kernel of FWHM 4mm and FILM prewhitening was used. To combine data within participant we ran fixed-effects analysis with cluster correction ( $Z > 2.3, p < .05$ ), we then defined LO and pFs using a method partially based on that proposed by Julian et al. (2012).

First, significant activation within each participant was binarised and linearly transformed into standard (MNI 152 T1 2mm) space. To identify the ‘average’ activation, the data were summed, spatially smoothed (Gaussian filter with FWHM 4mm), then divided by the number of participants (twelve). It was then thresholded at 0.6 to identify voxels where 60% of participants show significant activation. The thresholded activation in each hemisphere was then manually bisected into LO and pFs masks, based primarily upon anatomical location. These masks were then back-transformed into each participant’s individual space. Finally, for each of the left and right hemisphere LO and pFs masks in each participant we selected all active voxels lying within a sphere (10mm radius) centred on that participant’s peak voxel within the respective region. Using this method, we ensured that all participants had approximately the same number of voxels in their left and right LO and pFs ROIs (desirable for the Multi-Voxel Pattern Analysis (MVPA) described below), plus their ROIs were all in the same approximate anatomical location. This approach also has the secondary advantage of reducing overlap between LO-2 and the posterior parts of LO (see Table 5.1), allowing conclusions to be drawn for each region independently.

**Table 5.1.** Percentage overlap between LO-1, LO-2 and LO Localiser data across all participants

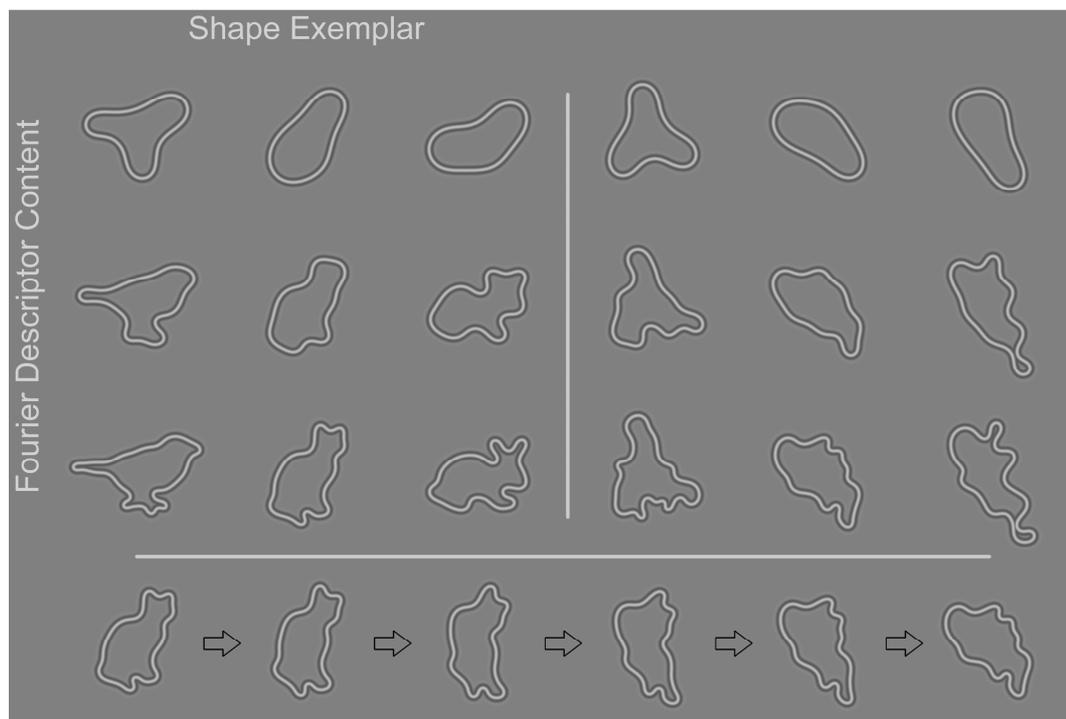
Participant	Sphere LO ROI		Cluster Corrected		Voxel Corrected	
	LO-1	LO-2	LO-1	LO-2	LO-1	LO-2
P01	0	2.1	0.5	19.3	0	6.2
P02	0	2.7	56.3	99.5	35.4	98.7
P03	0	0	3.0	16.4	0.8	12.4
P04	0	0	9.0	61.0	3.5	30.2
P05	0	13.3	38.8	95.5	8.9	87.5
P06	0.4	5.6	44.6	89.6	31.3	74.8
P07	4.3	22.8	33.3	70.0	22.4	49.3
P08	0	0.4	8.5	62.9	1.9	40.0
P09	0	0	31.5	56.4	7.9	38.4
P10	0	0	15.5	86.6	5.0	75.5
P11	0	18.1	16.8	75.5	5.0	57.3
P12	0	0	0	2.1	0	0.1
<b>Average</b>	<b>2.4</b>	<b>9.3</b>	<b>23.4</b>	<b>61.2</b>	<b>12.2</b>	<b>47.5</b>
<b>SD</b>	<b>2.8</b>	<b>8.8</b>	<b>18.5</b>	<b>32.6</b>	<b>12.7</b>	<b>32.3</b>

We calculated the percentage of voxels in LO-1 and LO-2 that overlapped with LO, or LOC localiser activity (Objects > Scrambled Objects) under various conditions. The first ‘Sphere LO ROI’ is the final LO ROI used for the purposes of this study, created using a sphere (10mm radius) centered on peak voxel (see Materials and Methods, LOC Localiser Scans). The second condition, ‘Cluster Corrected’, compares the percentage of voxels in LO-1 and LO-2 to the underlying activity on which the sphere ROI was created (i.e. cluster corrected LOC localiser data, Z-score thresholded at  $Z = 2.3$ ,  $p < .05$ ). The final condition, ‘Voxel Corrected’, is an additional analysis for comparative purposes. We thresholded LOC localiser data on a voxel-by-voxel basis ( $p < 10^{-5}$ , uncorrected for multiple comparisons), a common approach in the literature, and again compared percentage overlap with LO-1 and LO-2 voxels.

### 5.3.3 Main Functional Scans: Stimuli

To investigate the cortical representations of shapes, we developed a (standard) stimulus set comprising three exemplars (animal outlines), taken from the set of Snodgrass and Vanderwart (1980) images that had been converted to silhouettes and rated for recognisability by De Winter and Wagemans (2004). The exemplars were filtered on the basis of their Fourier descriptor (FD) content (Zahn and Roskies,

1972) to give three levels of detail (Low, Mid, High), yielding nine stimuli (Figure 5.2). More ‘detailed’ shapes contain more frequency information, and so they have greater variation in curvature around their perimeters. As such this can also be thought of as a manipulation in ‘curvature complexity’. By altering the phase of one FD (see methods below), we also created ‘scrambled’ versions to remove any semantic associations, which have been raised as a potential confound in previous literature (e.g. Haushofer et al., 2008b). These ‘scrambled’ stimuli were unrecognisable but matched to the ‘standard’ stimuli in FD content (Figure 5.2).



**Figure 5.2.** Stimuli used in the study. The ‘standard’ and ‘scrambled’ stimuli are shown on the top left and top right, respectively. Each column represents one exemplar, each row represents one level of Fourier Descriptor Content (low, mid, high detail; see Materials and Methods). The scrambled stimuli were created by rotating the phase of the Fourier Descriptor with most power through  $90^\circ$  counter-clockwise, as shown for one stimulus (in incremental phase rotations of  $18^\circ$ ) along the bottom row of the figure.

We calculated the set of FDs for a given image using the following procedure. The outermost boundary of the shape was extracted and moving average smoothing was applied, correcting for pixelation (for further detail see Chapter 2, Methods, section 2.8). The resultant smoothed boundary was then interpolated using a periodic cubic spline curve to 4,096 points. For every point around the contour, we then calculated both the distance and angle to the next point, before normalizing distance to the

range  $0-2\pi$  and removing the linear trend in the angles. By performing Fourier analysis on the set of angles, we can create our set of FDs. Critically, we save out information to make this process completely reversible, therefore shapes can be manipulated in the Fourier domain and transformed back to view the results. We removed linear trends from the resultant x- and y-coordinates to prevent non-closed boundaries.

To select three exemplars from the set of Snodgrass and Vanderwart (1980) images, we identified shapes that (1) exhibited a smooth, exponential decay of ‘power’ as a function of the number of FDs (as estimated by the norm of the residuals from a fitted exponential decay curve), (2) had high recognisability (ratings from De Winter and Wagemans, 2004) and (3) had biologically plausible shapes with relatively distinct profiles (i.e. animals). Using these criteria, we selected three shapes, specifically a bird, a cat and a rabbit, for each of these stimuli we then identified a ‘target FD’, which was a lenient estimate of the number of FDs needed to accurately reproduce the shape.

We next aimed to render our stimuli at low, mid and high levels of detail. The level of shape detail was manipulated by using a filter in the Fourier domain based on a Gaussian with a half width at half maximum (HWHM) that controlled the FD content of the shape. The filter was originally specified to have a maximum height of 2, but was subsequently clipped such that its value was unity for all values that originally exceeded 1.

For the three detail levels, we chose not to match the stimuli in terms of the number of FDs, as the number of FDs needed to describe a shape can vary somewhat arbitrarily. Instead, we matched detail across stimuli in terms of relative amplitude. First, for our three shapes we calculated the summed amplitude of the FDs (after normalizing it with the DC component) from 1 FD to the target FD defined above, then fitted a curve of form  $y=a(1-e^{bx})$  to the data (x: HWHM of Gaussian filter; y: sum of amplitude spectrum). We took the summed amplitude using our target FD and from our fitted curve, interpolated the HWHM needed to get  $\frac{1}{4}$ ,  $\frac{1}{2}$  and  $\frac{3}{4}$  of this value; creating low, mid and high detail boundaries (specifically, the HWHM values needed were 3.09, 2.45, 2.48; 7.16, 5.60, 5.85; 13.14, 10.02, 11.14 for the low, mid and high detail bird, cat and rabbit respectively).

The above process was used to create our standard stimuli. For the scrambled stimuli we performed the exact same procedure, except that the phase of the FD with most power in the low complexity shapes was rotated anti-clockwise through  $90^\circ$  (Figure 5.2). For the cat and rabbit, this was FD 2, for the bird it was FD 3. This manipulation meant that at low complexity the shapes largely changed through a rotation, whilst at high complexity interactions with higher frequency FDs caused our shapes to be completely unrecognisable. Critically, these scrambled stimuli share the exact same FDs (albeit with one phase change) as our standard stimuli.

The area of each stimulus was matched to the area of a square with length 6 degrees of visual angle. The profile of each shape outline was then rendered as the 4<sup>th</sup> derivative of a Gaussian (Wilkinson et al., 1998) at 50% contrast, yielding a peak spatial frequency of 1.68 cycles per degree.

#### **5.3.4 Main Functional Scans: Data Acquisition and Analysis**

Both the standard and scrambled functional scan sessions comprised five 8.5min stimulus presentation runs (TR = 2000ms, TE = 30ms, voxel size =  $2 \times 2 \times 2.5 \text{mm}^3$ , flip angle =  $90^\circ$ , matrix size  $96 \times 96 \times 26$ , FOV = 19.2cm). We employed a rapid event-related design, in which the stimulus presentation order had been counter-balanced and optimised (jittered) using Optseq2 (<https://surfer.nmr.mgh.harvard.edu/optseq>). Five stimulus presentation sequences were generated, one per run, and were used in order of most-to-least efficient (to ensure most efficient runs were used when participants were most alert). Stimulus presentation order was identical between the standard and scrambled scans, and 108 stimuli were presented per run (12 of each stimulus). Stimuli were presented centrally (centred on shape centroid) for 0.6sec with a median inter-stimulus interval of 3.4sec (interquartile range 2.4-6.4sec, range 1.4-18.4sec).

Participants maintained fixation on a red cross (0.60 deg. visual angle) and performed a one-back task (there were on average 10.6 sequential repeats per run). We prepended 10sec to the start of each scan to allow the magnetisation to reach a steady state, plus for all but the first two participant's standard scan sessions we appended 20sec to the end of the scan. This ensured that we were capturing the complete hemodynamic response for the final stimuli presented. During each scan,

we recorded video of the participant's right eye and later extracted eye blinks using custom written software.

Data were analysed using FEAT (Worsley, 2001). The first five volumes of each run were discarded (ensuring the scanner had reached stable magnetisation), the high-pass filter cut-off point was set to 100sec (correcting for low frequency drift), FILM prewhitening was used and motion was corrected for. Motion parameters were also entered as confound covariates. As in Op de Beeck et al. (2008b) we applied spatial smoothing (Gaussian kernel with FWHM of 4mm; twice voxel size). All nine stimuli were entered as separate explanatory variables (EVs), blinks were also added as an EV (modelled as a 200ms event from start of blink) as they can be a potential source of noise (Hupé et al., 2012; Gouws et al., 2014). Contrasts were set up to compare each individual stimulus to baseline, plus one contrast comparing the activity of all stimuli over baseline. To combine runs within participant, we ran fixed-effects analysis using cluster correction ( $Z > 2.3$ ,  $p < .05$ ). All data were retained in the high-resolution structural space.

Percent signal change was calculated on a voxel-by-voxel basis based upon the methods suggested by Mumford (2007).

### 5.3.5 ROI Restriction

All ROIs were restricted based upon the contrast of all stimuli over baseline from the functional scans of the complementary study (i.e. analysis using the standard stimulus set will use ROIs constrained using activity from the scrambled stimulus set). This was primarily performed because our stimuli only occupied a relatively small part of the visual field. As such, we would expect a smaller proportion of voxels in earlier visual areas such as V1 (with small receptive field sizes) to respond to our stimuli, in contrast to later areas such as LO-2 with receptive field sizes that respond to much larger regions of the visual field (Dumoulin and Wandell, 2008; Amano et al., 2009); constraining ROIs therefore helps to control for such differences. We used the activation from the opposite study to constrain ROIs as the stimuli in both studies occupy roughly the same spatial extent. This method also avoids concerns of circularity (Kriegeskorte et al., 2009) and does not cause the loss of power inherent in alternative approaches (e.g. split-half analysis).

### 5.3.6 Correlation Analysis and Shape Similarity Metrics

Our main (correlation) analysis explored the patterns of activity elicited by our various shape stimuli, and attempted to predict those patterns. We first created neural similarity matrices that captured for a given ROI the similarity (in terms of the pattern of neural activity) between all pairwise combinations of our shape stimuli. To achieve this, we took the whole-brain activation elicited by each of our nine stimuli (in units of percent signal change) and then for every ROI collapsed (concatenated) across hemisphere (V1-V4, LO-1, LO-2, LO, pFs) we extracted the pattern of activity specific to that area. We then iterated over all (36) pairwise combinations of our shapes and correlated the patterns of activity in each ROI independently (i.e. for each pairwise combination we correlated the activity elicited by the two shapes in that pair). In this way, higher correlations indicate which sets of stimuli elicited more similar patterns of activity in each of our ROIs.

To explore *why* certain stimuli elicit more similar patterns of activity in each region, the neural similarity matrices were then further correlated with a variety of different stimulus similarity measures, split into three broad categories. First, a perceptual measure was captured in a behavioural session completed post-scan (described below). Second, we had three low-level measures intended to capture ‘retinotopic’ shape similarity, these metrics depended highly on the exact location of our stimuli’s contours. The low-level predictors were:

- (1) Pixel-wise distance between the grey-values in pairs of our images. Whilst often a relatively crude metric, given that our shapes have been matched for size and translation, this should nevertheless capture some aspect of image overlap.
- (2) GIST descriptors (Oliva and Torralba, 2001). Each image is convolved with set of Gabor filters at 8 different orientations and 4 different spatial frequencies; these are split and averaged with a 4x4 grid. This results in 512 (8x4x16) unique values per image, which should capture the shape’s lower-level visual properties.
- (3) A contour discrepancy measure, essentially Procrustes analysis without matching for rotation. The ‘raw’ contours for a given pair of shapes were matched (in a least squares sense) for scale and translation (as this is how our final stimuli were presented to participants), we then took the average distance between all corresponding coordinates around those contours.

Finally, we had four more ‘abstract’ measures that were intended to capture the curvature in our shapes (e.g. number and magnitude of protrusions) irrespective of their spatial location. These were:

- (1) The number of minima and maxima (or concavities and convexities) around the shape’s contours.
- (2) The number of Fourier descriptors (or specifically, the HWHM of the Gaussian filter) needed to create our shapes; this is our most direct proxy for level of shape detail.
- (3) Shape compactness. This was represented with the area of the shape over the area of a circle with the same perimeter as that shape. Under this definition, a circle is (intuitively) the most compact shape and any deviations from circularity should decrease compactness.
- (4) A ‘convex perimeter’ measure. This was the perimeter of the shape over the perimeter of the shape’s convex hull. A convex hull is the smallest convex boundary that completely encapsulates the shape it is being fitted to. As such, concavities in the shape’s profile should increase the shape perimeter, but not necessarily the convex hull perimeter.

For all predictors, the final similarity metric for a given pair (Shape1-Shape2) was the Euclidean distance between the value for Shape1 and the value for Shape2. For measures returning multiple values (e.g. pixel-wise distance), the metric was the *average* Euclidean distance between the values for Shape1 and for Shape2. This metric was then inverted (subtracted from zero) so that larger numbers would represent greater similarity (i.e. zero represents perfect similarity).

### 5.3.7 Perceptual Similarity Measure

To acquire perceptual similarity ratings for our shapes, all participants performed two behavioural sessions (one per stimulus set) at least one week after the corresponding functional scan. These were primarily included as Op de Beeck, Torfs and Wagemans (2008b) found perceptual similarity predicted both LO and pFs activity whilst Haushofer, Livingstone and Kanwisher (2008b) found perceptual similarity only predicted activity in pFs. Our behavioural components therefore aimed to address the ambiguity regarding the role of perceptual shape similarity for neural representations in LO and pFs.

The stimuli used were identical to those described above (e.g. in size, position etc.), except that they were rendered upon a 400x400 pixel background of noise (i.e. every pixel was randomly set to a value between 0-255). We also generated 100 noise masks. The stimuli and masks were presented in a circular aperture that smoothed out to mid-grey at the edges (using a Gaussian filter with FWHM set to 90% of the circle's diameter).

The participant's task was to rate pairs of stimuli on a 1-6 Likert scale (extremely, very, slightly dissimilar to slightly, very, extremely similar). Stimulus pairs were presented with the following timings: noise mask (200ms), Shape1 (50ms), noise mask (500ms), Shape2 (50ms), noise mask (200ms). The experiment was split into four main blocks. The first block was a practice trial in which all (36) pairwise combinations of shapes were presented once, to familiarise participants with the task and comparisons. The second block contained four sets, and the third and fourth blocks contained three sets of all pairwise stimuli respectively. The experiment paused between blocks to provide a rest interval (participants pressed 'space' when they were ready to continue). Within each set of pairwise stimuli, the ordering was random, the pair ordering alternated such that if Shape1-Shape2 was a comparison in set one, set two would compare Shape2-Shape1. No comparisons were made between identical stimuli. Each behavioural session lasted approximately 30 minutes.

### **5.3.8 Multidimensional Scaling (MDS)**

In addition to the correlation approach, we also used multidimensional scaling on the neural similarity matrices from all ROIs across both stimulus sets. This allows us to visualise what each ROI's 'shape space' may look like. This MDS approach used PROXSCAL (Busing et al., 1997) in SPSS (IBM SPSS Statistics 20), with a weighted Euclidean or INDSCAL (Individual Differences Scaling; Carroll and Chang, 1970) model. This assumes a common dimensionality across all participants but allows for individual variability through use of weightings (i.e. one participant may preferentially weight dimension 1 over dimension 2, whilst another participant may do the converse, but ultimately both participants use the same dimensions).

Whilst the MDS was primarily used for visualisation purposes, we nevertheless aimed to 'quantify' the resultant solutions; two different approaches were used to

achieve this. First, we used a brute-force (stimulus blind) clustering approach. Specifically, for every ROI we iterated over all possible permutations of the nine stimuli, chunking each permutation into three sets of three items. The average inter-item distance was calculated within and then across each set, and the permutation that minimised this value was taken as the clustering solution. Essentially this method simply identifies the ‘best fitting’ set of three clusters (each containing three items) for all MDS solutions.

Our second approach specifically aimed to assess whether our stimulus dimensions (shape exemplar, FD-Content) were present in the extracted solutions. This analysis is analogous to the first, except that we just took the two stimulus permutations that clustered stimuli either by exemplar (bird, cat, rabbit) or level of detail (low, mid, high). For each ROI, the average inter-item distance for both permutations was calculated (as above) and then subtracted from the average distance across all pairwise combinations of shapes. If the result of this calculation was positive (i.e. greater than zero) then it would indicate some degree of clustering.

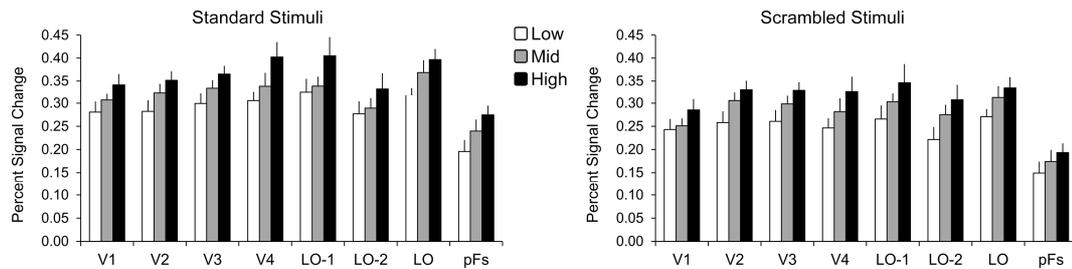
### 5.3.9 Statistical Analysis

All correlations were transformed to Fisher Z-Scores for averaging and statistical testing. For ANOVAs, the Greenhouse-Geisser correction was used when the assumption of sphericity had been violated (as indicated by Mauchly’s test) and corrected degrees of freedom are reported. All post-hoc tests are Bonferroni corrected.

## 5.4 Results

As an initial approach, we assessed percent signal change, as shown in Figure 5.3, across the low, mid and high detailed shapes, using ROI x FD-Content repeated measures ANOVAs.

Both standard and scrambled stimulus sets showed significant main effects of ROI (Standard:  $F(7,77) = 4.74, p = 1.8 \times 10^{-4}$ ; Scrambled:  $F(7,77) = 8.11, p = 2.3 \times 10^{-7}$ ) and FD-Content (Standard:  $F(2,22) = 9.79, p = .001$ ; Scrambled:  $F(2,22) = 12.54, p = 2.3 \times 10^{-4}$ ). There was no significant interaction in the standard stimuli ( $F(14,154) = 1.15, p = .318$ ) but a significant interaction emerged in the scrambled stimulus set ( $F(14,154) = 2.00, p = .021$ ).



**Figure 5.3.** Percent signal change for different brain regions. Results are shown for the standard (left) and scrambled (right) stimuli, when comparing Fourier Descriptor Content (FD-Content) collapsed across exemplar. Error bars indicate the standard error of the mean.

To explore the significant main effects, we ran Bonferroni-correct post-hoc tests across all ROIs and FD-Content levels. This revealed significantly diminished activity in pFs when compared to V3, V4, LO-1 and LO (Standard  $p = .005$ ,  $p = .022$ ,  $p = .039$  &  $p = .001$  respectively; Scrambled  $p = .003$ ,  $p = .006$ ,  $p = .003$  &  $p = 4.6 \times 10^{-4}$  respectively), plus V1 and V2 in the scrambled stimulus set ( $p = .049$ ,  $p = 9.8 \times 10^{-5}$  respectively). No other differences were significant (all  $p > .089$ ). For FD-Content, high detailed shapes elicited significantly greater activity than low detailed shapes (Standard  $p = .002$ ; Scrambled  $p = .002$ ), but there were no significant differences between low and mid (Standard  $p = .221$ ; Scrambled  $p = .080$ ) or mid and high (Standard  $p = .136$ ; Scrambled  $p = .056$ ) detailed shapes.

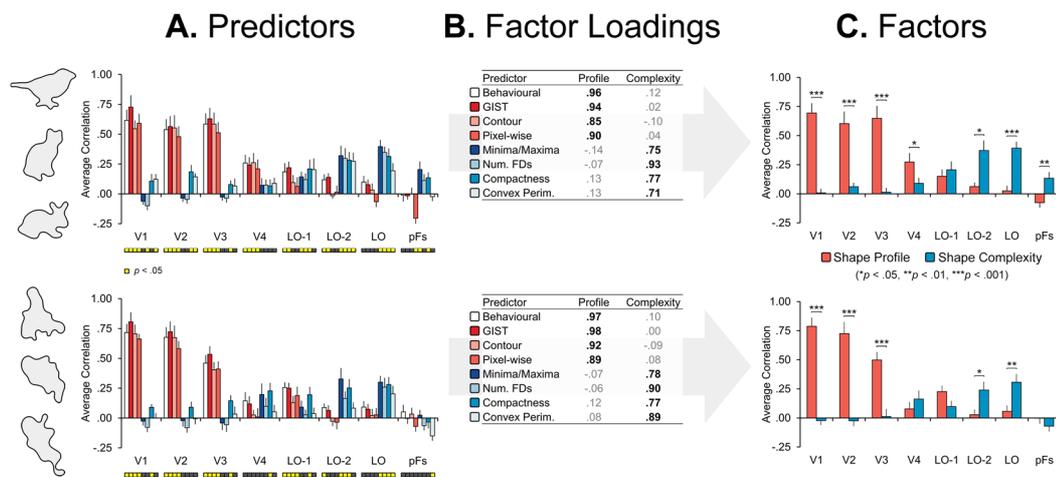
Given the significant interaction in the scrambled stimulus set's analysis, we ran one-way repeated measures ANOVAs comparing FD-Content within each ROI. All ROIs showed significant main effects of FD-Content (all  $F(2,22) > 5.27$ , all  $p < .014$ ), so Bonferroni-corrected post-hoc tests were used to compare across FD-Content levels. High detailed shapes elicited significantly greater activity than low detailed shapes in all ROIs (all  $p < .042$ ), plus significantly greater activity than mid detailed shapes in V4 ( $p = .012$ ) and LO-1 ( $p = .018$ ). Mid detailed shapes elicited significantly greater activity than low detailed shapes in LO-2 ( $p = .025$ ) and broadly V2 ( $p = .051$ ) plus V3 ( $p = .052$ ). No other comparisons emerged as significant (all  $p > .068$ ).

#### 5.4.1 Correlation Analysis

We next turn to our main analysis, taking the patterns of neural activity in each ROI and correlating them with our perceptual, low-level and more abstract similarity predictors. We reasoned that if a predictor correlated well with neural similarity

within a given ROI, then that predictor should inform us about the nature of the shape representation within that region.

The correlations between the similarity in activations and predictors for our ROIs are shown in Figure 5.4A. In general, we found that results for the perceptual measure were very similar to results for the low-level measures, and these in turn were strong predictors of activity patterns within V1-V3 and to a lesser extent V4. Conversely, our abstract measures were strong predictors of LO-2 and LO activity. In LO-1 neither low-level nor abstract measures appeared to dominate.



**Figure 5.4.** Correlations between predictors, factors and neural similarity. The figure shows from left to right: (A) the correlations between our predictors and neural similarity, for illustrative purposes bars that differ significantly from zero (two-tailed one-sample t-tests, uncorrected for multiple comparisons) are highlighted below each graph, (B) the loadings of these predictors in the two factors (Shape-Profile, Shape-Complexity) derived from principle component analysis and then (C) the correlations between those factors and the neural similarity matrices. The top and bottom rows of panels correspond to data obtained with standard and scrambled stimuli respectively. The middle tables also provide a colour key to the predictors plotted in the left hand panels. In the right most panel significance is derived from (uncorrected) two-tailed paired samples t-tests. Error bars indicate the (conventional) standard error of the mean.

Prompted by the pattern of results, we asked whether our predictors could be reduced to a smaller number of dimensions. We therefore used principal component analysis (PCA) on the predictors with orthogonal (varimax) rotation, and found for both stimulus sets two well-defined (independent) dimensions emerged (see ‘Factor Loadings’ in Figure 5.4B). A first ‘Shape-profile’ factor (Standard: 42.70%;

Scrambled: 44.94% variance explained) clearly captured the perceptual and low-level measures. This appeared to be characterizing the general spatial overlap between our shape outlines. A second ‘Shape-complexity’ factor (Standard: 31.77%; Scrambled: 35.25% variance explained) captured our more abstract image metrics linked to the curvature and FD-Content present in the shape outlines. We chose not to include a third factor as the first two factors already accounted for 74.48% (Standard stimuli) and 80.19% (Scrambled stimuli) of the variance; a third factor would have accounted for little additional variance (Standard: 11.68%, Scrambled: 9.92% variance explained). Furthermore, only the first two factors in both stimulus sets had eigenvalues greater than 1 in accordance with Kaiser’s (1960) criterion, and Horn’s (1965) Parallel Analysis also suggested the retention of just two factors for both stimulus sets.

We then correlated these Shape-profile and Shape-complexity factors with neural similarity for all ROIs (Figure 5.4C). This produced a clear pattern of results, with the Shape-profile factor dominating in early visual cortex but losing prominence in later extrastriate regions, where the Shape-complexity factor takes over. This pattern both highlights and clarifies the general trends noted earlier in the individual predictor correlations depicted in Figure 5.4A. The resultant correlations were evaluated with ROI x Factor ANOVAs for each stimulus set. For both stimulus sets, significant main effects of ROI (Standard:  $F(7,77) = 13.20, p = 4.7 \times 10^{-11}$ ; Scrambled:  $F(7,77) = 20.93, p = 1.7 \times 10^{-15}$ ) and Factor (Standard:  $F(1,11) = 15.43, p = .002$ ; Scrambled:  $F(1,11) = 40.21, p = 5.5 \times 10^{-5}$ ) emerged. However, these are qualified by highly significant interactions (Standard:  $F(2.80,30.80) = 35.24, p = 7.7 \times 10^{-10}$ ; Scrambled:  $F(7,77) = 36.83, p = 4.4 \times 10^{-22}$ ), making interpretation of main effects difficult. As such, we ran paired-samples t-tests within each ROI to compare factors.

For both standard and scrambled stimuli, the Shape-profile factor correlations were significantly greater than those for the Shape-complexity factor in V1, V2 and V3 (Standard:  $p = 1.0 \times 10^{-6}, p = 1.5 \times 10^{-4}$  &  $p = 1.1 \times 10^{-5}$  respectively; Scrambled:  $p = 1.0 \times 10^{-7}, p = 3.0 \times 10^{-5}$  &  $p = 2.0 \times 10^{-4}$  respectively). V4 showed a Shape-profile preference for the standard ( $p = .048$ ) but not scrambled ( $p = .361$ ) stimuli.

Conversely, we found significant Shape-complexity preferences in LO-2 and LO (Standard:  $p = .011$ ,  $p = 6.2 \times 10^{-5}$  respectively; Scrambled:  $p = .031$ ,  $p = .004$  respectively), plus pFs for the standard ( $p = .006$ ) but not scrambled ( $p = .293$ ) stimuli. In LO-1 we found no significant differences between the Shape-profile and Shape-complexity factors (Standard:  $p = .591$ ; Scrambled:  $p = .073$ ), this could either suggest that we failed to capture variance in LO-1, or that both factors play some role in describing its activity. To explore whether the correlations were greater than zero, we ran one-sample t-tests that found significant results for both the Shape-profile (Standard:  $p = .014$ ; Scrambled:  $p = 5.1 \times 10^{-4}$ ) and Shape-complexity (Standard:  $p = .009$ ; Scrambled:  $p = .041$ ) factors. This implies that the hypothesised retinotopic-to-functional transition could be occurring here.

#### 5.4.2 Result Reliability

To verify the robustness of our findings, we assessed alternative explanations and potential issues that might confound interpretation.

First, we aimed to address whether the shifting neural representation from V1-V3 to LO-1, LO-2 and LO is specific to these ROIs, or is it a more general property of neural tunings when moving anteriorly through visual cortex (perhaps due to increasing receptive field sizes for example)? To test this, we analysed two additional retinotopically-defined areas, V3A/B (anterior and dorsal to V3d) and combined regions VO-1 and VO-2 (VO-1/2) (Brewer et al., 2005) that extend anteriorly from V4 (the VO-1/VO-2 boundary was not always clear across participants hence they were collapsed to a single ROI). These ROIs were again constrained with the ‘all stimuli over baseline’ contrasts (as per the main study). For the standard stimuli, the average correlation between the Shape-profile and Shape-complexity factors with neural similarity in V3A/B were  $r = .30$  and  $r = .07$ , respectively, and they differed significantly ( $t(11) = 3.65$ ,  $p = .004$ ). A similar profile of results was found for VO-1/2 (Shape-profile  $r = .20$ ; Shape-complexity  $r = .00$ ;  $t(11) = 2.86$ ,  $p = .015$ ). For the scrambled stimuli the corresponding V3A/B correlations were  $r = .16$  and  $r = .14$ , which in this case were not significantly different ( $t(11) = 0.19$ ,  $p = .854$ ). In VO-1/2 there was no evidence that either factor was represented (Shape-profile  $r = .05$ ; Shape-complexity  $r = .02$ ;  $t(11) = 0.52$ ,  $p = .612$ ). Whilst the correlations with scrambled stimuli were generally weaker, the fact that with standard stimuli the Shape-profile correlation was significantly greater

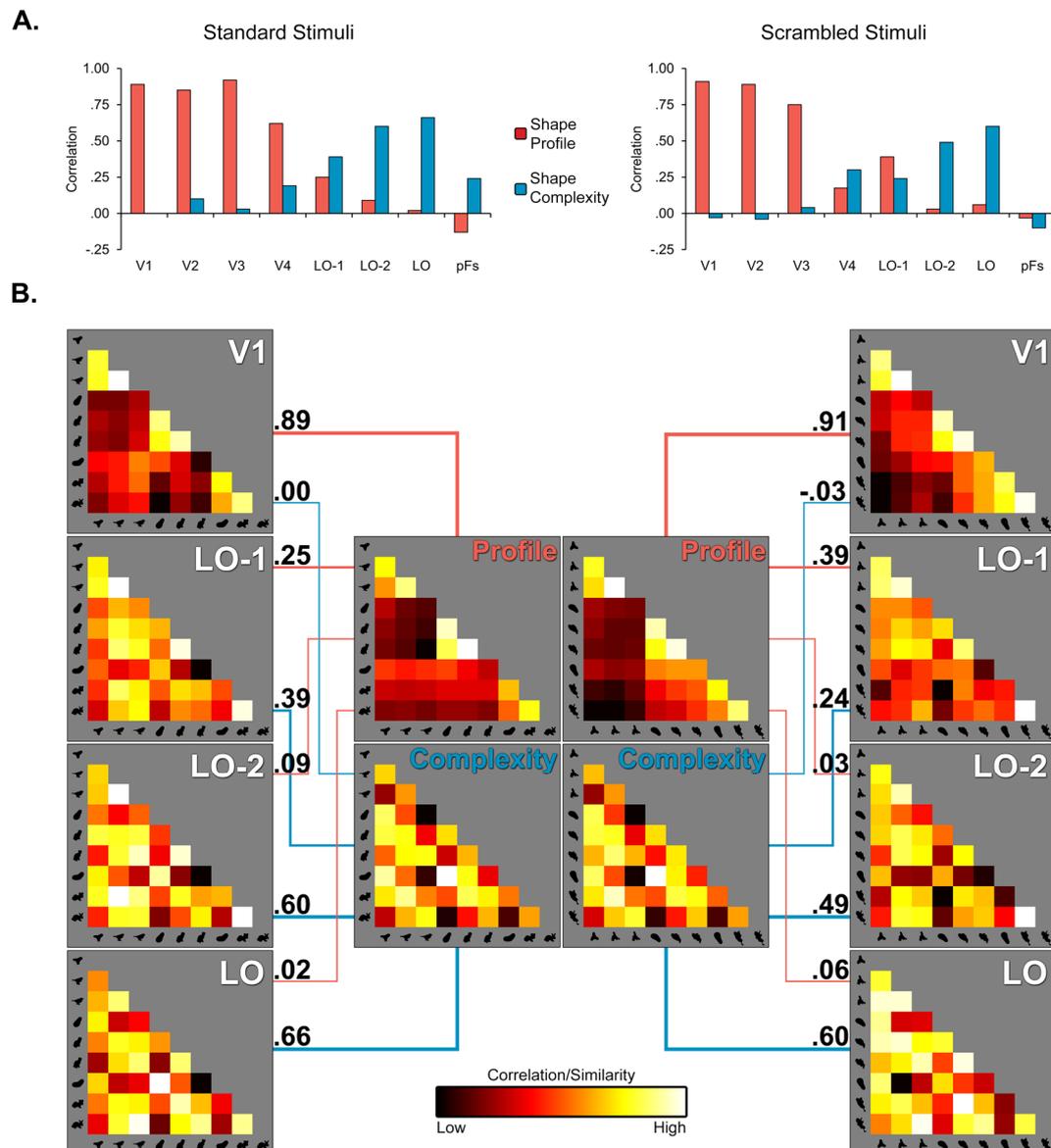
than the Shape-complexity correlation in both V3A/B and VO-1/2 demonstrates that more anterior ROIs do not necessarily transition to the more abstract/curvature-tuned representation that emerged around LO-1 and LO-2.

Second, a potential issue is the use of spatial smoothing. It is possible that smoothing blurred the boundaries between ROIs and perhaps results such as LO-2's Shape-complexity preference could be explained by spread from neighbouring LO. To test this, we re-ran all analysis without spatial smoothing and correlated the Shape-profile and Shape-complexity factors with the resultant neural similarity measures across all ROIs. In general, very small differences were observed; with average losses of 0.77% (Standard stimuli) and 1.93% (Scrambled stimuli) variance explained when spatial smoothing was not used. As a lenient test for differences, we then ran two-tailed paired-samples t-tests to compare correlations in the original analysis to those produced without spatial smoothing. The results reported in Table 5.2 suggest that generally spatial smoothing was beneficial, as correlations in larger ROIs that you would expect to be less susceptible to spatial smoothing (e.g. V1 or LO) were significantly reduced without it. In contrast, for a smaller ROI (LO-1) we observe a significant increase in the Shape-complexity correlation, implying a lack of spatial smoothing was actually beneficial. Importantly, the interpretation of the data analysis remains the same whether or not spatial smoothing was used.

**Table 5.2.** Comparing shape factor/neural similarity correlations with and without spatial smoothing

	V1	V2	V3	V4	LO-1	LO-2	LO	pFs
<b>Standard Stimuli</b>								
Shape-profile	.69/.71	.60/.63	.65/.60	.27/.24	.15/.10	<b>.06/.02*</b>	<b>.02/-.01*</b>	-.08/-.05
Shape-complexity	.01/-.01	.06/.06	.01/.03	.09/.06	.20/.23	.37/.32	.39/.33	<b>.13/.05*</b>
<b>Scrambled Stimuli</b>								
Shape-profile	<b>.79/.76*</b>	<b>.73/.68*</b>	<b>.50/.37**</b>	.08/.06	.23/.19	.03/.05	<b>.06/-.01**</b>	.00/-.03
Shape-complexity	<b>-.02/-.05*</b>	-.02/-.02	.01/.04	.16/.13	<b>.10/.13**</b>	.24/.20	<b>.31/.21*</b>	-.07/-.06
Format: With spatial smoothing/Without spatial smoothing					* $p < .050$ , ** $p < .010$ , *** $p < .001$			

Exploring the impact of spatial smoothing on correlational data. Spatial smoothing (Gaussian kernel with FWHM 4mm) was used when analysing our main functional scans, potentially blurring representations across smaller ROIs. To test whether this was an issue, we compared our main correlation results (see Figure 5.4C) with and without spatial smoothing applied (significance derived from uncorrected two-tailed paired-samples t-tests). Generally, a lack of spatial smoothing appears to reduce correlations slightly, however the drop in variance explained is minimal. As such spatial smoothing does not appear to be a confound in our analysis.



**Figure 5.5.** Group mean neural similarity matrices and their correlation with ‘Shape-Profile’ and ‘Shape-Complexity’ factors. (A) Correlations between standard (left) and scrambled (right) group mean similarity matrices and the Shape-Profile and Shape-Complexity factors. (B) Neural similarity matrices for the standard (left) and scrambled (right) stimuli in V1, LO-1, LO-2 and LO (collapsed across participant), plus factor similarity matrices centrally. Here ‘brighter’ colours represent greater similarity; note that to enhance visibility the colour map has been scaled independently and linearly in each image based upon the minimum and maximum correlations, maximizing the range of colours used in each image. As with our main correlation analysis, we see V1 corresponds closely to the Shape-Profile factor whereas LO-2 and LO show greater similarity with the Shape-Complexity factor. LO-1 is linked to both factors, with the exact balance changing slightly between stimulus sets.

Finally, the main correlation analysis examined how our factors predict neural similarity on an individual basis, however this contains individual variability as a source of noise. To provide a cleaner picture of our results, we therefore collapsed neural similarity across participants and compared the resultant correlation matrices with our factors. No differences emerged in the pattern of correlations, although as expected on the basis of reducing noise our factors now captured more variance (Figure 5.5A). Collapsing neural similarity across participants also allows us to visualise our results. Figure 5.5B depicts for both stimulus sets the averaged *neural* similarity matrix between all pairwise combinations of the shape stimuli in four key ROIs, as well as the pairwise *stimulus* similarity described by the two factors (Shape-profile, Shape-complexity).

This approach provides some insight into the organisational principles underlying our Shape-profile and Shape-complexity factors. For example, the similarity between the ‘mid detail bird’ and ‘mid detail rabbit’ (column 2, row 2 from bottom left in each similarity matrix) is very low in the Shape-profile factor due to the minimal spatial overlap, however they are highly similar in the Shape-complexity factor. This implies that our Shape-complexity factor captures some index of shape curvature or complexity not grounded in retinotopic coordinates.

We also note that the similarity matrices in Figure 5.5B look markedly similar across studies. To quantify this, we correlated the V1, LO-1, LO-2 and LO average similarity matrices from the standard stimuli with those from the scrambled stimuli; all showed highly significant correlations (V1:  $r = .62$ ,  $p = 4.7 \times 10^{-5}$ ; LO-1:  $r = .74$ ,  $p = 2.0 \times 10^{-7}$ ; LO-2:  $r = .79$ ,  $p = 1.1 \times 10^{-8}$ ; LO:  $r = .81$ ,  $p = 2.3 \times 10^{-9}$ ). If we instead compare the equivalent correlations across individuals, the results are naturally lower (V1:  $r = .33$ ; LO-1:  $r = .22$ ; LO-2:  $r = .29$ ; LO:  $r = .18$ ), but two-tailed one-sample t-tests show that nevertheless they are all significantly greater than zero (V1:  $p = .001$ ; LO-1:  $p = .029$ ; LO-2:  $p = .003$ ; LO:  $p = .025$ ).

These results first demonstrate the degree to which averaging similarity matrices across participants can reduce individual variability as a source of noise, particularly in later extrastriate regions such as LO. More importantly, they show that the fundamental relationships between the stimuli within a set are largely preserved after ‘scrambling’, at least in terms of the characteristics that appear to be influential

across visual cortex. This is perhaps not as surprising for V1, as similarity is driven by spatial overlap so (for example) the low, mid and high detail cats would all be highly similar. As the same FD is rotated for all cat exemplars during the scrambling process, the outlines will change in similar manners and so spatial relationships are maintained. It is more notable for the Lateral Occipital representations, as the results imply that scrambling also preserves the curvature complexity relationships across exemplars. This means that the scrambled stimuli are good controls for the more recognisable counterparts.

### 5.4.3 Multidimensional Scaling (MDS)

Our final exploratory approach, multidimensional scaling, used the neural similarity matrices from a given ROI to test for underlying structure in terms of how that ROI responds to shape stimuli (see MDS solutions for representative ROIs in Figure 5.6A). Only two-dimensional solutions were extracted as this allows for easy visualisation, allowing us to perceptually assess the resultant solutions and providing some idea of how each ROI may be representing the shape stimuli. Crucially, whilst our correlation analysis could only explore neural similarity with respect to our factors (based on stimulus properties), MDS takes no account of the relations between stimuli as it is based on neural data alone.

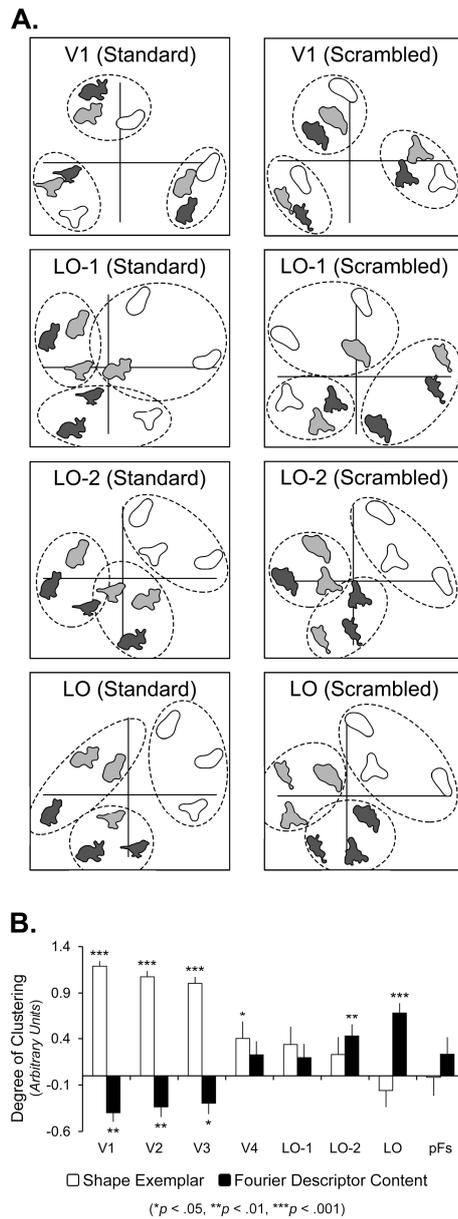
We first ran the ‘stimulus blind’ clustering analysis on our representative ROIs to provide some insight into the underlying structure. Specially, for each ROI we identified the optimal 3x3 grouping (i.e. three groups of three items) that minimised the average pairwise distance between the items within each group (see Materials & Methods). The clusters (see circled items in Figure 5.6A) indicated that in V1 items with a similar shape profile were grouped together, whilst the clusters in later extrastriate areas (namely LO-2 and LO) were largely grouped on the basis of shape/curvature complexity. No consistent clustering emerged in LO-1.

Next, we performed analysis on the MDS solutions across all ROIs to specifically determine whether one or both of our original stimulus dimensions (from Figure 5.2) were present in the resultant solutions (i.e. do shapes cluster on stimulus exemplar; bird, cat or rabbit, or Fourier Descriptor content; low, mid and high detail). Note that we are not exploring how our factors (Shape-profile, Shape-complexity) relate to the MDS solutions as MDS essentially just re-describes the neural similarity data (albeit

with reduced dimensionality); the main correlation analysis described above has already explored the factor/neural similarity relationship in the ‘purest’ sense.

No significant differences in clustering patterns emerged between the standard and scrambled stimulus sets (paired-samples t-tests, all  $p > .173$ ) and so for simplicity we collapsed across stimuli. The clustering analysis in Figure 5.6B again reveals a shift in the nature of shape representations between early and late visual cortex. Shape exemplar is most influential in early areas such as V1, whereas FD-Content (or level of shape detail) predominates in later extrastriate regions LO-2 and LO. Quantifying this, one-sample t-tests indicated that the shape exemplar dimension showed significant clustering in V1 ( $p = 9.1 \times 10^{-14}$ ), V2 ( $p = 4.5 \times 10^{-12}$ ), V3 ( $p = 5.2 \times 10^{-11}$ ) and V4 ( $p = .041$ ). In fact, for V1, V2 and V3 the FD-Content dimension showed significant negative clustering (i.e. dispersion; V1:  $p = .001$ ; V2:  $p = .008$ ; V3:  $p = .022$ ), likely due to the tight exemplar clusters driving down the average pairwise distance. Only LO-2 and LO showed significant clustering on the FD-Content dimension (LO-2:  $p = .003$ ; LO:  $p = 5.7 \times 10^{-6}$ ). No dimensions in LO-1 or pFs differed significantly from zero (all  $p > .099$ ), implying a lack of clustering.

Generally, these results support our main findings. Early retinotopic regions are dominated by the spatial layout of a shape’s outline, whereas Lateral Occipital ROIs (namely LO-2/LO) are dominated by more abstract measures (likely the size and number of protrusions around each shape). In LO-1 neither dimension appears to dominate, so it is a viable candidate for the transition point between these two organisational principles. Interestingly, LO-2 appeared to broadly split the shape exemplar and FD-Content dimensions orthogonally (Figure 5.6B), implying sensitivity to shape profile that was not captured in our factor analysis (although hints emerged in the raw predictors, see behavioural and GIST predictors; Figure 5.4A). This would fit with the retinotopic organisation of the area, however our clustering analysis did show that the FD-Content dimension nevertheless dominates (in line with the main analysis).



**Figure 5.6.** Solutions from multidimensional scaling. (A) Extracted two-dimensional solutions for four representative brain regions in the standard and scrambled stimulus sets. For FD Content, Low, Mid and High detailed shapes have been shaded in white, mid- and dark-grey respectively. Dashed rings highlight the tightest set of three clusters (each containing three items) for each ROI (see Materials and Methods). Here we see strong shape exemplar clustering in V1, a general lack of clustering in LO-1, then evidence for FD Content clustering in LO-2 and (to a greater extent) LO. Also of note, for LO-2 there is some evidence of an orthogonal separation between FD Content and shape exemplar. The dispersion accounted for each ROI MDS solution is (Standard stimuli/Scrambled stimuli): V1 .93/.95, LO-1 .82/.86, LO-2 .86/.86, LO .89/.86

(B) Graphical depiction of clustering based upon the stimulus dimensions (Figure 5.2); here we plot the mean distance between all stimuli minus the distance between all pairs within a given stimulus dimension. A value of zero implies clustering within a dimension is no greater than the average clustering across all pairwise combinations of items, whereas positive values imply some degree of clustering. Significance values are taken from one-sample t-tests assessing whether the clustering differs significantly from zero. Error bars indicate standard error of the mean.

## 5.5 Discussion

A general pattern of results emerged across our studies using standard and scrambled stimuli, with activity in lower-level retinotopic ROIs (V1-V3) being better predicted by a shape's spatial layout, whereas extrastriate Lateral Occipital regions LO-2 and LO were better predicted by more abstract shape features. A middle ground emerged

in LO-1, which could represent a transitional point between these two organisational principles. The more abstract representation appears to be specific to Lateral Occipital cortex, as more abstract tunings were not found in ventral ROIs V4 and VO-1/2, or the more dorsal ROI V3A/B.

First, we note that very similar results were identified across both stimulus sets, despite their considerable differences, demonstrating the findings' robustness. It also reinforces the lack of semantic influences both in early visual areas and in LO, in line with previous work (Kim et al., 2009). This places LO as perhaps an earlier shape-processing region, one that encodes the form of the shape regardless of its novelty, recognisability or familiarity.

The results from a standard GLM analysis across ROIs identified a generally flat profile of responses to our shape stimuli, with the exception of pFs which showed diminished activity. This highlights the importance of exploring representations using multivariate approaches, given that total signal change revealed very few differences along the visual hierarchy. We also noted consistently greater activation for high complexity shapes (with slight differences between areas), likely due to the increased amount of detail in their contours.

Turning to our main findings, both our correlation and MDS results implied that there were (at least) two general organisational principles in visual cortex. First, we have 'low-level' organisation ('Shape-profile'), in which the precise layout of the shape drives activation (i.e. translation, scaling or rotation would alter similarity measures under this definition). Unsurprisingly, this measure was a strong predictor for early retinotopic cortex. Secondly, we have a more abstract organisation ('Shape-complexity') that appears to be capturing curvature in the shape's profile (translation, scaling, rotation would not alter 'abstract' similarity measures). This measure dominated in LO.

Of key interest is how these two organisational principles relate to LO-1 and LO-2. LO-1 appeared to contain influences of both low-level and abstract representations, whereas LO-2 showed remarkable similarity to LO's results profile. In effect, we seem to have a retinotopic ROI responding non-retinotopically. However, receptive field sizes increase progressively through the visual hierarchy (Dumoulin and Wandell, 2008; Amano et al., 2009), implying LO-1 and (to a greater extent) LO-2

will be pooling information over larger areas compared to earlier regions. Larger receptive field sizes could allow for a more lenient response profile to visual stimuli, meaning regions like LO-2 can respond to more abstract features (such as curvature-complexity) instead of the strict, Shape-profile based activation of areas with smaller receptive fields. However, it is worth noting that our stimuli were presented in a relatively limited region of the visual field, perhaps biasing tunings in LO-2 away from a ‘retinotopic’ response profile. It is likely that our low-level measures of shape overlap would play a greater role in predicting neural similarity if stimulus spatial position was an experimental manipulation, particularly given evidence of sensitivity to spatial position in LO (Grill-Spector et al., 1999; Sayres and Grill-Spector, 2008; Schwarzlose et al., 2008). Nevertheless, spatial jitter may have masked the more subtle shape representations that emerged in this study, therefore controlling for stimulus position was likely beneficial.

Given that LO-2 in particular shows some overlap with LO (Sayres and Grill-Spector, 2008; Silson et al., 2016), is causally involved in shape-processing tasks (Silson et al., 2013) and both LO-2 and LO appear to share the same abstracted representations, there is strong, converging evidence that this region is a preliminary object-processing stage as Larsson and Heeger (2006) posited. However, whilst it is possible that this more abstract representation emerges solely within LO-2, its complete lack of Shape-profile tuning implies that there may be some prior transitional point where more abstract tunings start to emerge. A good candidate for the transition is LO-1, in which patterns of activity could be explained (albeit weakly) by both low-level and more abstract stimulus properties. As such, whilst the functions of LO-1 and LO-2 are clearly dissociable to some extent (Silson et al., 2013), a stream of activation passing from LO-1 to LO-2, then LO (and beyond) seems plausible. This has a second implication; that LO could contain more explicit retinotopy than the general lower visual-field biases that have been previously identified (Sayres and Grill-Spector, 2008), indeed retinotopic maps have been reported in this area (Kolster et al., 2010). Whilst we have clearly found that a more abstract dimension predominates in this area, the same is true for LO-2 which does contain retinotopic maps. As such, an ‘abstract’ representation does not preclude underlying retinotopy.

The nature of the ‘abstract’ representation emerging in LO-2 and LO is perhaps less intuitive than the concept of low-level physical similarity. Our ‘abstract’ similarity measures were clearly proxies for some underlying construct, given that they reduced so cleanly to a single dimension, and this dimension appears to be capturing the contours, or more specifically the curvature of the shape. This finding is in accordance with Macaque literature demonstrating the importance of curvature information for the ventral visual stream (Kayaert et al., 2011; Yue et al., 2014). It also corroborates Drucker and Aguirre (2009), who found amplitude manipulations in composite radial frequency patterns (making protrusions more or less salient) were linked to LO neural similarity. However, we also found that the *number* of curves in a shape’s profile strongly predicted similarity in LO-2 and LO (see minima/maxima; Figure 5.4A), implying *both* amplitude and frequency are important. Work in a subsequent chapter (Chapter 7) will evaluate how these two variables underpin shape representations in LO. In terms of future work, it would also be valuable to probe the respective roles of convexity and concavity. It is well-established that convexities are generally highly salient visual features (Bertamini and Wagemans, 2013) and convexities elicit greater activation than concavities (Haushofer et al., 2008a). Moreover, LO can represent shapes when parts of the contours are deleted or obscured (Kourtzi and Kanwisher, 2001; Lerner et al., 2002; Stanley and Rubin, 2003; Hayworth and Biederman, 2006), suggesting a complete contour is not needed to identify shapes. Convexities and concavities must occur together, and so fully representing both could lead to redundancy. The broader evidence therefore points to an underlying representation based upon shape convexity.

Finally, it is intriguing that the abstract shape features failed to reliably predict neural activity within pFs. This region was less active in comparison with other areas, so perhaps responses were not robust enough to establish clear neural similarity measures. Alternatively, individual differences in perceived shape similarity could underlie this discrepancy. Our behavioural measure was originally intended to capture such between-subject differences, however it appears as though participants were predominantly using spatial overlap to determine perceptual shape similarity, given that the behavioural measure collapsed into the same dimension as the low-level predictors. This meant that our metric for perceptual similarity did not correlate

with LO or pFs activity in contrast to previous work (Haushofer et al., 2008b; Op de Beeck et al., 2008b). Haushofer et al. in particular noted that responses within pFs (but not LO) reflected a more ‘implicit’ metric of shape similarity (how likely two different shapes were to be confused as identical), which is perhaps better suited to measuring subtle perceptual differences than explicitly asking for similarity ratings. Such perceptual differences may underlie the unexplained variance in pFs activity in our study.

To conclude, we identified two orthogonal organisational principles in visual cortex; Shape-profile and Shape-complexity. Whilst the Shape-profile factor reflected the spatial layout of our shape stimuli, the Shape-complexity factor instead reflected changes in curvature around the shapes perimeter. Accordingly, our Shape-profile factor was a strong predictor for early retinotopic cortex, whereas the Shape-complexity factor correlated well with shape-selective LO. Critically, LO-1 and LO-2 (two retinotopic regions lying between early visual cortex and later extrastriate regions) show intermediate representations. LO-1 has influences from both factors whereas in LO-2 the Shape-complexity factor dominates. We argue that this represents a transitional point between two previously discrete approaches; retinotopy and functional selectivity.

## **Chapter 6. Testing for Linearity in Fourier Descriptor ‘Information’ Content**

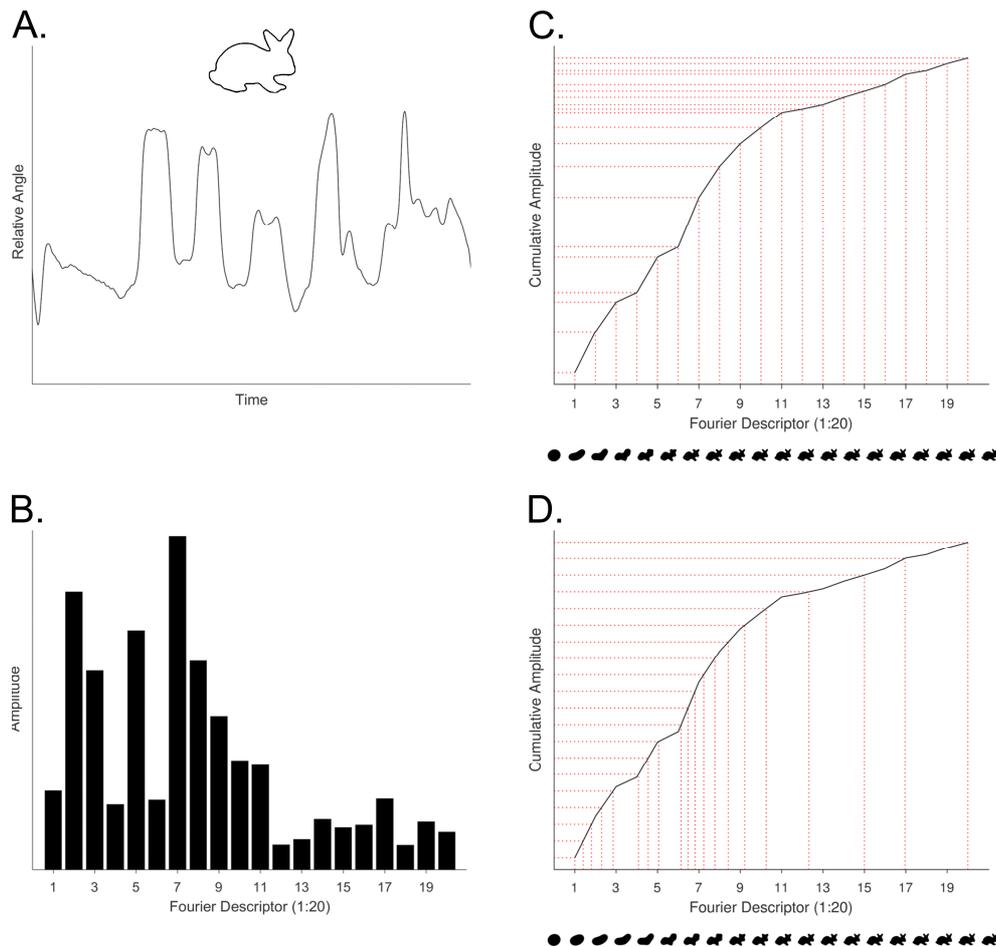
### **6.1 Abstract**

This experiment aimed to test whether participants were able to make accurate midpoint judgements on shapes that had been manipulated in terms of summed Fourier amplitude content. We used the psychophysical method of bisection to determine sets of stimuli that participants judged to be equidistantly spaced. We then tested the linearity of those judgements by comparing them to a variety of predictors based upon the physical characteristics of our stimuli. Summed amplitude content proved to be the best predictor of participants’ responses. This suggests that summed amplitude may capture perceptually salient details in a shape’s contour, and that manipulations in summed amplitude are a useful means to probe underlying shape representations.

## 6.2 Introduction

In the previous chapter (Vernon et al., 2016) our design involved two assumptions. First, we assumed that linear increases in the cumulative sum of Fourier Descriptor (FD) (Zahn and Roskies, 1972) amplitude would generate a better stimulus set than linear increases in the number of FDs (Figure 6.1). This was because the number of FDs needed to render a given shape varies considerably, and the level of detail each FD adds also depends heavily on the nature of the base shape. For example, consider a set of radial frequency patterns (Wilkinson et al., 1998); the amplitude of a given FD will depend almost entirely upon the frequency of the current pattern (i.e. a three lobed pattern will have high amplitude at FD 3). In contrast, we reasoned that the cumulative sum of amplitudes should be a better measure of shape complexity, as it captures additional information regardless of the frequency at which that information is contained. Manipulations in cumulative amplitude can be achieved as rather than selecting a discrete number of FDs, we filter the Fourier spectrum with a trimmed Gaussian (see section 5.3.3 in the previous chapter for further details). This allows us to render a shape using partial FDs (analogous to non-integer FDs), allowing much more flexibility in choice of shape contours, and therefore level of detail.

Our second assumption was that the scrambled shapes would act as reasonable controls to the original recognisable images. To scramble the shapes, we took the FD with greatest amplitude in the low-complexity shape and rotated its phase through  $90^\circ$  counter-clockwise. This renders the shape unrecognisable whilst keeping fixed the FD amplitude spectrum of the original shape; hopefully resulting in comparable shape complexity. As such, this assumption also rests on the perceptual equivalence in complexity of fixed FD spectra, albeit with slightly altered phase spectra.



**Figure 6.1.** Choice of linear transitions. To create a set of Fourier Descriptors (FDs), we take a shape and traverse around its contour, plotting relative angle over time (A; see Chapter 2, Methods section 2.8 for further details). We can then take the Fourier transform of this ‘time series’ and evaluate the amplitude spectrum (B; showing the first 20 FDs only). To create linear transitions in shape detail, there are two choices available. These choices are illustrated in C & D, which each depict the cumulative summed amplitude from B. First, we could increase linearly in the number of FDs (C), however this results in uneven increases in cumulative amplitude. Alternatively, we could increase linearly in cumulative amplitude (D), and whilst this leads to uneven increases in number of FDs, we reasoned that this method should provide smoother increases in contour complexity. Example shapes rendered at the 20 example FD values are shown beneath C and D respectively.

We aimed to test this equivalence by asking whether participants could make linear judgements in summed FD amplitude content during an unconstrained behavioural task. Specifically, we asked participants to manipulate a central shape flanked by two exemplars until it appeared to be perceptually midway between those exemplars

(method of bisection, e.g. Rensink and Baldridge, 2010). Critically, the two exemplars were lower and higher complexity than the central shape, and participants were manipulating the central shape by adjusting its summed amplitude content. If they could identify the ‘correct’ midpoint (in terms of amplitude content) of the central shape, then it would suggest that they were able to make linear judgements in such content. For example, if the lower and upper shapes corresponded to the 1<sup>st</sup> and 3<sup>rd</sup> horizontal lines in Figure 6.1D (i.e. rabbit exemplars 1 & 3), would their midpoint choice correspond to the 2<sup>nd</sup> horizontal line (i.e. rabbit exemplar 2)?

## 6.3 Methods

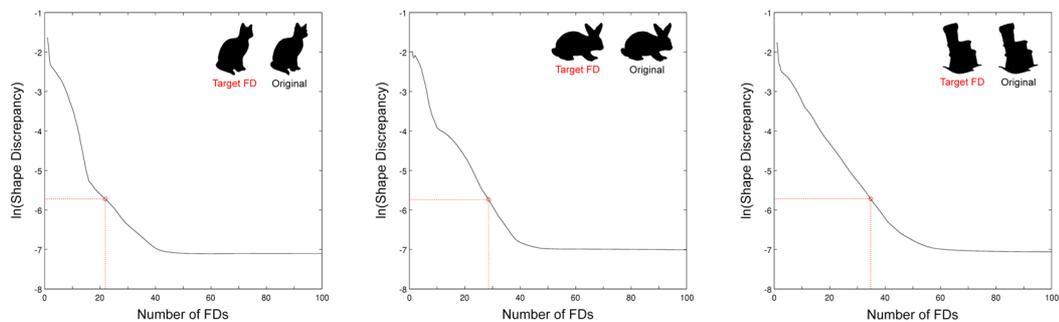
### 6.3.1 Participants

Twelve participants (8 females) were recruited from the University of York Psychology department. Half of the participants had participated in the study described in the previous chapter, and so were familiar with shape amplitude/complexity manipulations. The remaining participants were completely naïve with regards to the stimulus set.

### 6.3.2 Stimuli Creation

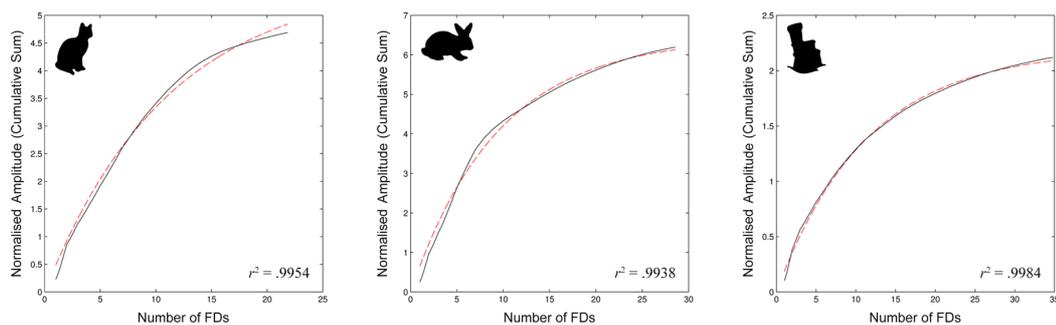
We aimed to create a stimulus set of 6 shapes, half of which would be recognisable, the other half scrambled as in the previous study. This should allow us to test how comparable recognisable and scrambled stimuli actually are. We also wanted half the stimuli to be from the previous study, and the other half entirely novel to explore effects of familiarity. With this in mind, we took the ‘cat’ and ‘rabbit’ shapes from the original study, as well as a novel ‘chair’ stimulus. To scramble, we again rotated the FD with greatest amplitude through 90° counter-clockwise, however we no longer had the constraint that the FD rotated needed to be present in a low-complexity exemplar. This meant that whilst for the cat (and also the chair), we still rotated FD 2, for the rabbit it was now FD 7 that was rotated (note the peak at FD 7 in Figure 6.1B), creating a different scrambled outline. This does mean that the rabbit and ‘scrambled rabbit’ are equivalent up to FD 7 (which can be seen later in Figure 6.4), but they rapidly diverge after that point. To summarise, the familiar shapes are the cat, rabbit and scrambled cat. The novel shapes are the chair, scrambled chair and scrambled rabbit.

To allow manipulations in FD spectral content, we now aimed to create a set of 200 exemplars for each stimulus that increased linearly in cumulative amplitude content. To explain how this was done, we first need to go into some additional detail regarding how complexity was controlled for in the previous experiment. First, for every shape we have a target FD that is an estimate of the number of FDs needed to (broadly) recreate the original shape. To estimate this, we took the shape rendered with 1 FD and 100 FDs, and then calculate a ‘shape discrepancy’ measure. This is the average Euclidean distance between the original contour and filtered contour, after matching via Procrustes analysis. Critically, if we take the natural log of this shape discrepancy measure, it shows a largely monotonic decrease in average discrepancy as the number of FDs used to render the contour increases (i.e. as the filtered shape becomes closer to the original shape), before plateauing at some point (Figure 6.2). We exploited this by performing a binary search to identify the number of FDs needed to get a shape discrepancy matching (to the nearest thousandth) the (arbitrarily chosen) 75<sup>th</sup> percentile on the linear range between 1 and 100 FDs. This number was the ‘target FD’, and represents a value where the shape is easily recognisable, without reaching the point of diminishing returns (i.e. when many more FDs are required to only slightly increase shape accuracy).



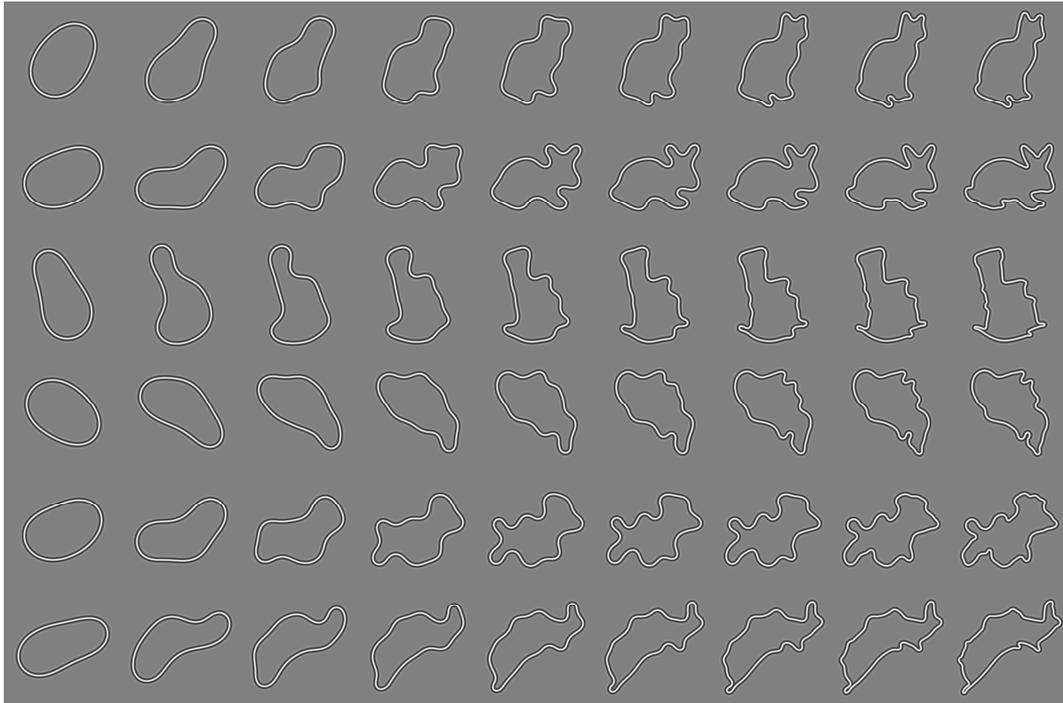
**Figure 6.2.** Target Fourier Descriptor (FD). By plotting natural log of shape discrepancy (degree contours differ) against number of FDs, we see a roughly linear decrease followed by a plateau. We take the 75<sup>th</sup> percentile on this line (dashed red line/circle) as the target FD, i.e. the maximum number of FDs we will use for a given shape. In the upper right corner of each graph we compare the shapes rendered at this target FD, to the original outline. There are minimal perceptible differences, highlighting the success of this approach. Note that shape discrepancies may appear small, this is because part of the shape matching process resizes image to have unit area.

The next step is to be able to predict the summed amplitude at a given number of FDs. To do this, we first render the shape from 1 FD to the target FD defined above, in 1,000 linearly spaced increments. The summed amplitude (after normalisation with the DC component) is then saved out at each level. Finally, we fit (in a least squares sense) an increasing exponential decay curve of form  $a(1-e^{-bx})$  to these data so that we can estimate the number of FDs required for a given level of summed amplitude (Figure 6.3). We now take the summed amplitude at the target FD and generate 200 linearly spaced points from ‘15% target FD amplitude’ to ‘target FD amplitude’. By using our fitted curve, we can then estimate the number of FDs needed to produce that summed amplitude for each of those 200 points and create our stimulus set. Essentially, for each of the 6 shapes we now have 200 exemplars from low to high complexity, that show a good approximation to linear increases in summed amplitude across the range.



**Figure 6.3.** Predicting FD amplitude. The black line shows actual summed amplitude (normalized by DC component) for 1,000 exemplars in each shape ranging from 1 FD to the target FD defined earlier. The dashed red line shows the best fitting curve of form  $a(1-e^{-bx})$ ; this allows us to predict the number of FDs needed for a given amplitude (avoiding the computationally intensive process of calculating linear steps manually).

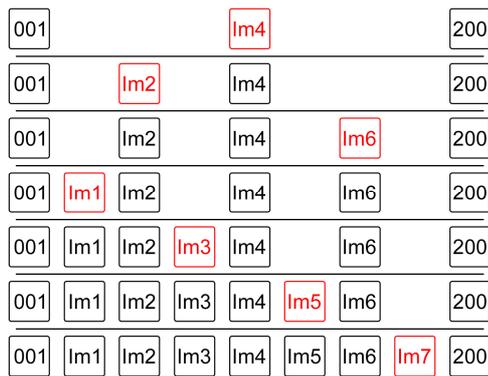
As in the previous study, the area of each stimulus was matched to the area of a square with length  $6^\circ$  visual angle. The profile of each shape outline was then rendered as the 4<sup>th</sup> derivative of a Gaussian (Wilkinson et al., 1998) at 50% contrast, yielding a peak spatial frequency of 1.68 cycles per degree (Figure 6.4).



**Figure 6.4.** Example stimuli. Here we show 9 broadly linearly spaced exemplars (in terms of summed FD amplitude) for each of our six shape stimuli (cat, rabbit, chair followed by corresponding scrambled counterparts). If participants make perfectly linear choices, the central 7 stimuli here will be their midpoint choices for the 7 comparisons (the 1<sup>st</sup> and 9<sup>th</sup> stimuli here are the fixed endpoints). Note that these stimuli are rendered at 100% contrast for visibility, not 50% contrast as used experimentally.

### 6.3.3 Stimulus Presentation

First, to give a broad overview of the paradigm; we aimed to test participant’s bisection judgements through 7 comparisons. In each comparison, the participant will aim to select the shape that bisects two flanking exemplars (i.e. the method of bisection). For example, the starting comparison will use exemplars 1 and 200 and if participants are making judgements based on a linear relationship with summed FD amplitude, they should adjust the shape until it lies around exemplar 101. The 7 comparisons go in the order specified in Figure 6.5.



**Figure 6.5.** Image comparison order.

Exemplars 001 and 200 are fixed and represent the endpoints from Figure 6.4, Im1:Im7 (i.e. Image 1 to Image 7) are selected by the participant. The current comparison is highlighted in red, and the comparisons go in the following order:

1. Im4 – bisecting exemplars 001 and 200
2. Im2 – bisecting exemplars 001 and Im4
3. Im6 – bisecting exemplars Im4 and 200
4. Im1 – bisecting exemplars 001 and Im2
5. Im3 – bisecting exemplars Im2 and Im4
6. Im5 – bisecting exemplars Im4 and Im6
7. Im7 – bisecting exemplars Im6 and 200

By arranging the exemplar choices for Im1 through to Im7, we can test to see if midpoint judgements are based on linear perception of summed amplitude content (i.e. a perfectly linear response would be to select exemplars 26, 51, 76, 101, 125, 150 & 175 for Im1-Im7 respectively; Figure 6.4).

The image to be adjusted was presented centrally (centred on shape centroid), set against a mid-grey (RGB 128 128 128) background on a gamma corrected CRT monitor (1600x1200 @ 85Hz). The comparison stimuli (flanking exemplars) were centred  $11.84^\circ$  to the left and right of the screen centre respectively (centred vertically). Participants adjusted the complexity of the central shape with the arrow keys on a standard keyboard; the up and down arrows incremented the shape by  $\pm 5$  exemplars, the left and right arrows incremented the shape by  $\pm 1$  exemplar to allow more fine-scaled tuning. The final choice was selected with the ‘space’ key. Complexity manipulations were limited to the range set by the two flanking comparison stimuli. The transitions were ‘smooth’, meaning that holding a key down would keep incrementing or decrementing shape complexity until the key was released (at a rate of 20Hz). This ensured that participants could not count the ‘correct’ interval between two exemplars and simply split it. For half the participants, the lower complexity exemplar was on the left, the higher complexity exemplar on the right. For the other half, this was reversed (counterbalanced across familiar and naïve participants), to control for congruency effects.

Each comparison was repeated 4 times, and a given comparison was completed for all 6 images (in pseudo-random order) before being repeated. This meant a given comparison and its repeat were generally well separated in time (identical sequential comparisons were not possible) so participants were unlikely to simply remember their choice for each repeat. Once the four repeats had been completed for a given comparison, the rounded average was taken as the final choice.

Participants performed the task in a dark room, resting on a chin-rest placed 57cm from the screen. They were instructed to make their best guess regarding the comparisons, and that there was no right or wrong answer. We simply asked them to try and make the central stimulus appear perceptually in the middle of the exemplars on either side. There was no time limit with regards to a comparison. In total, the behavioural session lasted approximately 30 minutes depending upon participant speed.

#### **6.3.4 Predicting linearity**

After participants perform their midpoint judgements, we aimed to determine what stimulus metric they were internally using to adjust the shapes. The most obvious choice would be predicted summed amplitude (based upon the fitted increasing exponential decay curve), as this is the manipulation that participants were actually making. However, if they fail to show linear judgements in amplitude then we may be able to find an alternative predictor that provides some insight into how judgements were being made. Alternatively, even if midpoint judgements reflected linear perception of summed amplitude, this could just be a proxy for some other metric which may better capture the recorded responses. As such, we created a set of predictors to see which would best predict the participants’ judgements. For comparative purposes, these were predominantly based upon the predictors utilised in the previous chapter (Vernon et al., 2016), so the relevant descriptions have been taken from that section.

The first set of predictors (after predicted summed amplitude) were ‘low-level’ shape overlap metrics:

1. GIST descriptors (Oliva and Torralba, 2001). Each image is convolved with set of Gabor filters at 8 different orientations and 4 different spatial frequencies; these are split and averaged with a 4x4 grid. This results in 512 (8x4x16) unique

values per image, which should capture the shape's lower-level visual properties. To compare two images, we take the average absolute difference between the set of 512 values for each image.

2. A contour discrepancy measure, essentially Procrustes analysis without matching for rotation. The 'raw' contours for a given pair of shapes were matched (in a least squares sense) for scale and translation, we then took the average discrepancy about those contours.
3. Average pixel-wise distance between the grey-values across two images.

These measures are all inherently comparative, that is, a comparison between two images is necessary to produce a (meaningful) single number for each individual image. For example, the average grey value in a given image will tell us little, but the average difference across grey values in two images may be informative.

Therefore, these predictors were generated by comparing all subsequent shapes (2:200) to shape 1.

The next set of predictors were based upon our more 'abstract' metrics from the previous study. These aimed to capture aspects of curvature complexity:

4. The number of minima and maxima (or concavities and convexities) around the shape's contours.
5. The number of Fourier descriptors (or specifically, the HWHM of the Gaussian filter) needed to create the shapes.
6. Shape compactness. This was represented with the area of the shape over the area of a circle with the same perimeter as that shape.
7. A 'convex perimeter' measure. This was the perimeter of the shape over the perimeter of the shape's convex hull.

These predictors are not inherently comparative; they each produce single, meaningful numbers for a given image and so could predict responses *as is*. Nevertheless, to ensure approaches were comparable across the low-level and abstract predictors, we also generated these values as the absolute difference between shapes 2:200 and shape 1. This should have no impact on correlation analyses however, as it would just shift all values by a set amount.

The final predictor to use is *actual* summed amplitude. To explain, whilst we could predict summed amplitude relatively accurately using our fitted exponential curves, the fit is not perfect (Figure 6.3). As such, once we have our set of shape exemplars that are perceived as linearly spaced (as positioned by participants) we can average the choices at each level and calculate actual summed amplitude for the chosen shapes. For example, if the average middle choice for the cat stimulus was exemplar 104, then we can recreate that stimulus, calculate its FDs and hence its actual summed amplitude, rather than simply relying on the predicted value.

### 6.3.5 Statistical Analysis

All correlations were transformed to Fisher Z-Scores for averaging and statistical testing. For ANOVAs, the Greenhouse-Geisser correction was used when the assumption of sphericity had been violated (as indicated by Mauchly’s test) and corrected degrees of freedom are reported.

## 6.4 Results

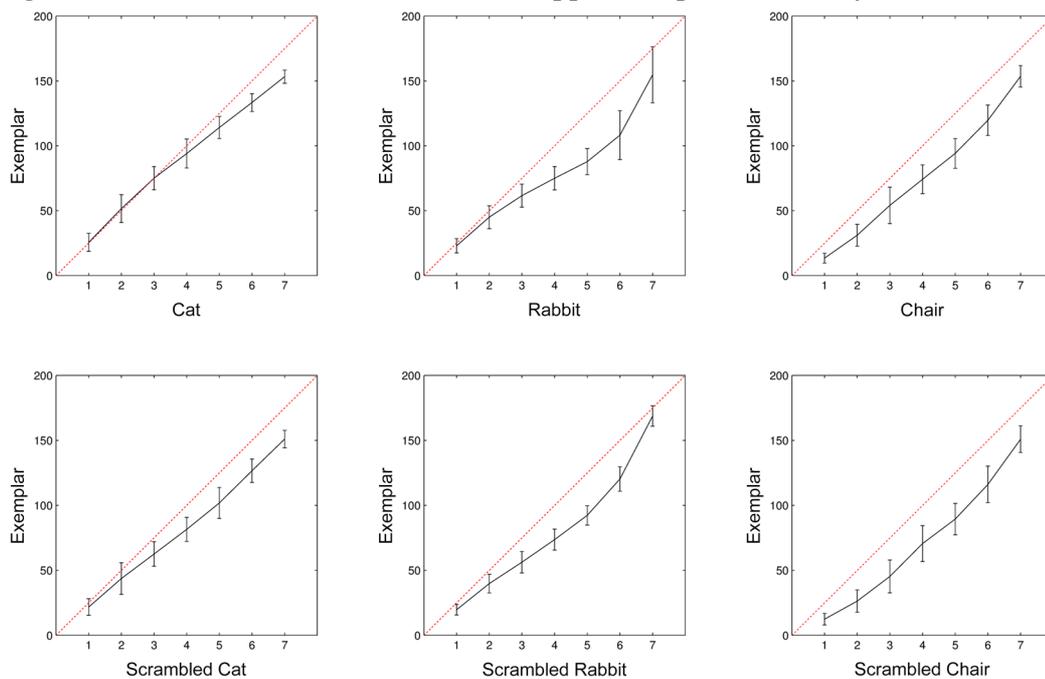
### 6.4.1 Counterbalancing Measures

We first tested whether our counterbalancing measures influenced results, by collapsing average choices across shape in each counterbalanced group, then testing for differences with non-parametric tests (as our data is not normally distributed). Mann-Whitney U tests indicated that participant familiarity (i.e. familiar or naïve with regards to shape manipulations) had no effect on ratings ( $U = 24.00, p = 1.00$ ) and neither did the positioning of exemplars (i.e. whether the low and high complexity endpoints were on left- and right-hand sides respectively, or vice versa) ( $U = 22.00, p = .805$ ).

We then tested for differences between recognisable and scrambled stimuli, plus new or old stimuli using Wilcoxon signed-ranks tests. Whilst there was no significant effect of scrambling ( $Z = 1.69, p = .109$ ), we did find that old stimuli had significantly higher general ‘scores’ than new stimuli ( $Z = 2.03, p = .047$ ). To check whether this could be a confound, we broke scores down across familiar and naïve participants (as the ‘old’ stimuli would only be old to familiar participants). Both sets of participants showed the exact same rankings (i.e. both showed significant old/new differences;  $Z = 2.03, p = .047$ ), suggesting that the difference is a property of the shapes themselves, and not due to shape novelty.

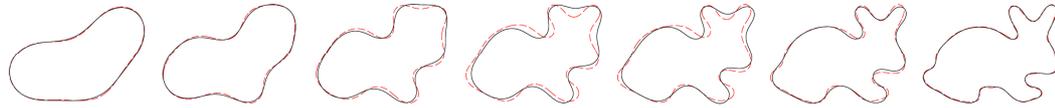
### 6.4.2 Stimuli Results

As we now have no evidence to suggest between-participant differences, the average scores for each shape across the 7 comparisons were collapsed across participants for visualisation purposes (Figure 6.6). There are a few things to note. First, both the chair and scrambled chair stimuli lie consistently below the linear line, likely explaining why new shapes elicited significantly lower scores on average than old shapes. In fact, none of the scores across stimuli lie exactly on the perfect linear line, however they do show broadly linear profiles. The fact that choices generally lie beneath or drop away from the ideal lines likely implies that diminishing returns had already taken effect by the 75<sup>th</sup> percentile of shape discrepancy (see Figure 6.2), and so our target FD was perhaps too lenient in some cases (i.e. perceptible contour detail likely plateaus before our target FD is reached). Supporting this, when we compare linear choices to those the participant actually made for one shape stimulus (Figure 6.7), we see that the linear choices appear to ‘peak’ too early. A more



**Figure 6.6.** Average participant linearity choices. Along each x-axis we show the 7 comparisons participant were asked to make, and along the y-axis we plot the exemplars chosen for each comparison. Exemplars are broadly linearly spaced in terms of summed amplitude content. The red dashed line indicates perfectly linear choices. Whilst participant’s choices do not necessarily lie on the perfect linear line, they do nevertheless appear broadly linear. Note that where deviations from linearity are observed, they are typically mirrored across recognisable and scrambled stimuli; this will be explored further in the main text. Error bars represent  $\pm 1$  SD.

stringent choice for target FD may have kept responses linear over a greater range. The choices may also reflect a slight bias towards lower frequencies, which makes sense given that these will have a more significant effect on the resultant shape outline.

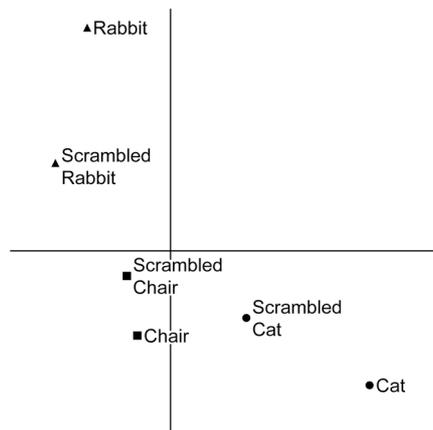


**Figure 6.7.** Participant versus Linear responses. Here, we compare (for one shape) the set of linear choices (red dashed lines) versus participants’ choices (black lines). We note that the linear choices perhaps approximate the true outline too early, indicating that a more constrained stimulus range may have been preferable. However, the shapes do match up towards the end of the stimulus set; this could instead just indicate a lower frequency bias (i.e. participants may be more sensitive to manipulations in the lower frequency range).

Intriguingly, where deviations from linearity are observed, they seem to be matched across the standard and scrambled stimuli. For example, the rabbit shows a ‘dip’ peaking around comparison 5 or 6 (Figure 6.6), and a similar dip is observed for the scrambled rabbit. To statistically test whether standard and scrambled stimuli were more similar to each other than would be expected by chance, we ran a correlation analysis. Specifically, we had a key set of ‘matched spectrum’ pairs (Cat – Scrambled Cat; Rabbit – Scrambled Rabbit; Chair – Scrambled Chair) and there are 14 other sets of possible pair combinations in our stimulus set. We correlated the pairs of responses in each pair combination, and then collapsed across pairs. The average correlation for our set of matched spectrum pairs was  $r = .995$ , whereas the average correlation across all other sets was  $r = .988$ . A one-way repeated measures ANOVA across the 15 (Fisher Z transformed) pair sets showed a highly significant effect of pair set ( $F(14,154) = 9.2 \times 10^{-15}$ ). Exploring this, planned comparisons indicated that the matched spectrum pair set had significantly higher correlations than all other pair sets (all  $p < .005$ ). Effectively, this confirms that participants’ midpoint judgements were more similar for recognisable stimuli and their scrambled counterparts than any other shape pair combination.

To visualise similarity, we also tried multidimensional scaling (MDS) on our data. Specifically, for every participant we correlated all pairwise combinations of the shapes, then took  $1 - r$  to turn results into a dissimilarity matrix. We then used the

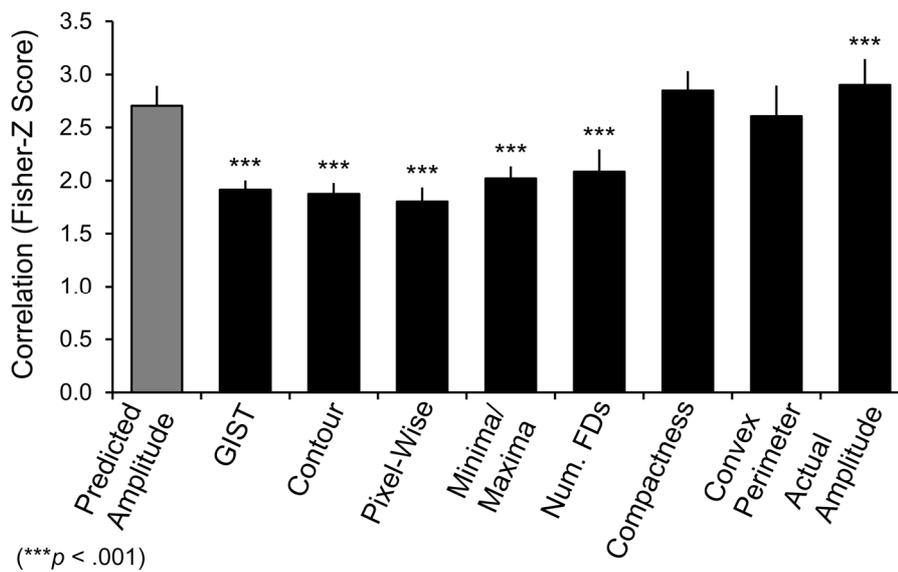
same MDS approach as in the previous chapter, specifically PROXSCAL (Busing et al., 1997) in SPSS (IBM SPSS Statistics 20), with a weighted Euclidean or INDSCAL (Individual Differences Scaling; Carroll and Chang, 1970) model. The results (Figure 6.8) again suggested that recognisable and scrambled stimuli were given relatively similar ratings.



**Figure 6.8.** Multidimensional Scaling solution for participant’s midpoint judgements. This plot shows similarity (on arbitrary axes) between our shape stimuli, in terms of the exemplars participants chose as linear (similarity is represented by distance; shorter distances imply greater similarity). Note that a given shape and its scrambled counterpart tend to be close together. This implies that on some level, recognisable and scrambled stimuli have a degree of similarity. The dispersion accounted for in the model was .872.

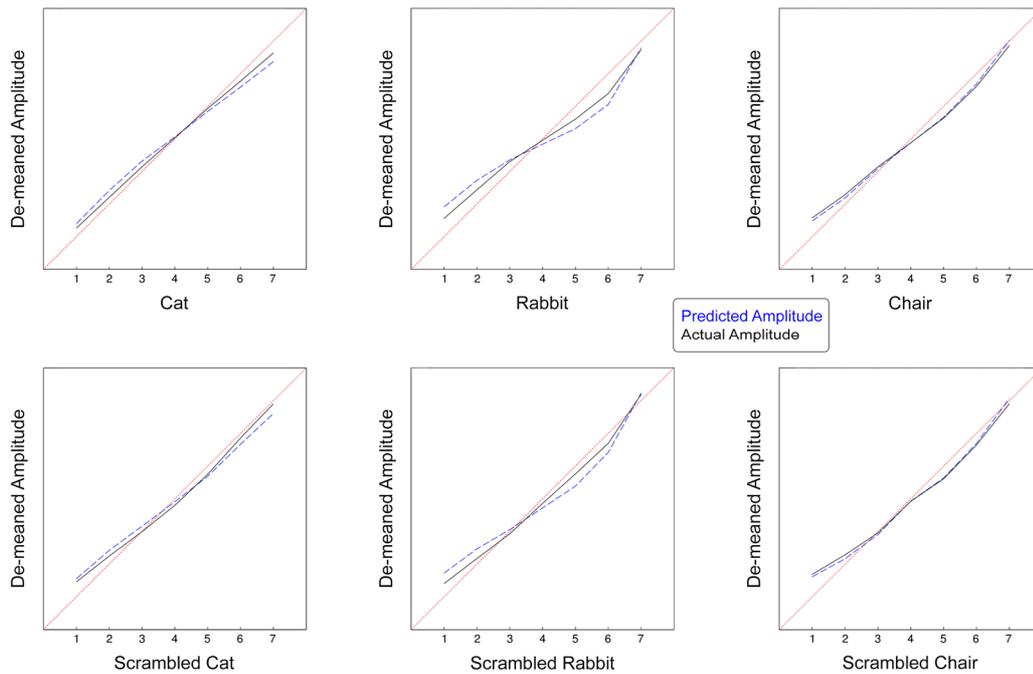
### 6.4.3 Predicting midpoint judgements

Finally, we tested to see which stimulus characteristics (as used in previous work) could best predict the choices participants chose as linear. Our null hypothesis is that predicted amplitude will best explain results, as that was the manipulation participants were actually making. Against this, we aim to test predictors from the previous study (GIST descriptors, contour discrepancy, pixel-wise differences, minima/maxima, number of FDs, compactness and the convex perimeter measure), plus the actual amplitude content of the shapes. A one-way repeated measures ANOVA showed a highly significant effect of predictor ( $F(1.66, 18.24) = 73.52$ ,  $p = 5.2 \times 10^{-9}$ ; Figure 6.9). Planned contrasts comparing ‘predicted amplitude’ against all others, showed that it was a significantly better predictor than all ‘low-level’ measures (GIST, contour, pixel-wise; all  $p < 2.6 \times 10^{-7}$ ), plus minima/maxima and number of FDs ( $p = 1.6 \times 10^{-6}$ ;  $p = 9.5 \times 10^{-8}$ ). There were no significant differences between compactness or convex perimeter when compared to predicted magnitude ( $p = .175$ ;  $p = .069$  respectively), although both measures were numerically larger. Intriguingly, actual amplitude scores were the best predictor out of the set, working significantly better than predicted amplitude ( $p = 2.8 \times 10^{-5}$ ) at predicting participant’s responses.



**Figure 6.9.** Comparing the predictive power of stimulus characteristics to explain participant midpoint judgements. All predictors are compared against ‘Predicted Amplitude’ as that was the metric participants were manipulating. We note that our more ‘abstract’, curvature-complexity based predictors (namely ‘Compactness’ & ‘Convex Perimeter’) are not significantly worse than ‘Predicted Amplitude’. However, only ‘Actual Amplitude’ significantly surpasses it. Error bars show 1 SD.

To visually evaluate the predicted versus actual amplitude difference, we plotted the relevant values across exemplar choice as shown in Figure 6.10. All values were de-meant, to account for general shifts in the vertical direction. As discussed earlier, such shifts could be a consequence of a lenient target FD, and so they are not necessarily relevant in regards to the question of whether participants were able to make judgements in summed amplitude. The plots corroborate the results from Figure 6.9; actual amplitude values tend to lie closer to linearity than predicted values (albeit to varying extents across stimuli, dependent upon the accuracy of the original estimates). This implies that participants were likely accounting for slight inaccuracies in our predicted amplitude values, suggesting that some aspect of Fourier amplitude content is capturing perceptually relevant details.



**Figure 6.10.** Predicted versus actual amplitude. Along the x-axis we plot the 7 exemplars participants chose as a linear progression. Along the y-axis we plot the corresponding cumulative amplitudes based upon our predictions (blue dashed line) as well as the actual values generated from the stimuli (black solid line). The red dotted line indicates a perfectly linear choice. All amplitude values have been de-meaned, to account for general vertical shifts. Note that the black line generally lies closer to true linear judgements, suggesting participants may have been partially accounting for prediction errors when choosing their midpoints.

## 6.5 Discussion

This study has been relatively self-contained, largely serving as a post-hoc verification for previous design choices. Namely, in the previous study we chose to manipulate summed amplitude rather than number of FDs, as intuitively that seemed like a more appropriate metric. The validity of our scrambled stimuli as controls also in part rested upon the salience of the amplitude spectrum. We checked these assumptions by testing whether participants could make midpoint judgements based on summed amplitude content, in an unconstrained behavioural paradigm (i.e. participants were free to make any choice they felt was appropriate, across a large range of exemplars).

Generally, we found that participants were able to make judgements in summed amplitude that were monotonically related to ‘predicted amplitude’ and were highly

linearly correlated. Indeed, ‘predicted amplitude’ was a significantly better predictor than all ‘low-level’ metrics, plus ‘number of FDs’, implying that it is a more perceptually relevant measure to vary stimuli on. It could be argued that because participants were manipulating shapes based upon predicted amplitude, our results are weighted towards finding that metric as the best predictor. Whilst likely true to some extent, this is countered by the fact that ‘actual amplitude’ was a significantly better predictor than ‘predicted amplitude’. This implies that participants were internally compensating for inaccuracies in the fits from Figure 6.3 (that aim to predict number of FDs for a given amplitude). That is, participants did not seem to be simply basing their judgements upon the linear range we had set, instead it seems they were basing judgements upon the true amplitude of our stimuli. Nevertheless, it would be interesting to create a linear scale based upon e.g. one of our lower-level predictors (such as GIST descriptors) to test whether the resultant midpoint judgements still produced linear ranges in terms of summed amplitude.

Of equal interest, was the fact that when participants deviated from linearity, these deviations occurred concomitantly across both recognisable and (matched spectrum) scrambled stimuli. Multiple statistical tests confirmed that responses to matched spectrum stimulus pairs were generally more similar to each other than any other set of shape pairs. This implies that our scrambled stimuli were good controls in the previous study. It also corroborates the idea that participants were basing responses on the amplitude spectrum, as that was the only common aspect across recognisable and scrambled stimuli.

However, we are still left with the question of exactly what participants are responding to. We do not suggest that human shape processing relies on the calculation of Fourier descriptors and summed amplitude. Some hints come in the form of our predictor performance. In the previous chapter we asked whether the Shape-complexity dimension (that captured variance across Lateral Occipital Cortex) was dependent upon the frequency or amplitude of our stimuli. Here, we note that the number of minima or maxima (analogous to frequency) is a comparatively poor predictor, hinting that amplitude may be more salient. Indeed, as amplitude increases the features within a shape will become more pronounced. This could be seen as a transition from obtuse curvature at low complexity towards acute curvature as complexity increases (see Figure 6.4; for example, the ‘obtuse’ head of the rabbit

gradually tends towards acute curvature as features, namely ears, are defined). This would explain why our compactness predictor did not significantly differ from predicted amplitude; compactness will be heavily influenced by acute curvature. There is evidence that shape-selective LOC is sensitive to amplitude information (Drucker and Aguirre, 2009), and this proposal would be in line with the wider literature suggesting a curvature-based shape representation in Macaque V4 (Pasupathy and Connor, 1999, 2001, 2002; Connor et al., 2007; Yue et al., 2014). As such, amplitude-based curvature sensitivity seems a plausible account of these data.

Still, participants’ midpoint judgements were not perfect and there were some notable deviations from linearity. We have already discussed one potential cause. It is possible our target FD (estimated number of FDs needed to approximate complete contour) was too lenient. This was set at a relatively arbitrary value (75<sup>th</sup> percentile of log shape discrepancy; Figure 6.2), aiming to balance contour complexity against diminishing returns (i.e. needing relatively large increases in amplitude for relatively small changes in contour detail). However, participants’ choices were either completely below the linear line, or generally dropped off as complexity increased. This implies that they were placing more weight on lower frequencies, perhaps due to smaller differences in the later end of the range. A lower target FD might have produced a more linear range in terms of participants’ responses.

An alternative possibility is that not all contour changes are equally salient. For example, in the rabbit example (Figure 6.7), the ears are ‘locked in’ relatively early in terms of linear changes, however for participants’ choices these features change more slowly. Selective attention to parts of the shape might affect the rate at which participants adjust the amplitude content, consequently eliciting deviations from linearity. It is also possible that we are assessing two independent processes here. As discussed, deviations from linearity tended to occur as complexity increases.

Therefore, there may be linear effects over the lower frequency range, with alternative influences emerging as complexity increases (perhaps feature salience starts to play a role here). Indeed, in the radial frequency (RF) pattern (Wilkinson et al., 1998) literature, there are noted differences between low frequency and high frequency RF shapes (e.g. Loffler et al., 2003). There is the suggestion that lower frequency shapes may be processed globally (as opposed to processing local features

such as curvature) (Loffler, 2008), and such mechanisms may be playing a role in the lower complexity range of our shape exemplars.

Finally, we should note that the results in this chapter, highlighting the importance of the amplitude spectrum, do not undermine the importance of the phase spectrum. In Chapter 2 (Methods, section 2.5.1), we demonstrated that the amplitude spectrum was relatively unimportant compared to the phase spectrum of an image’s 2D Fourier transform. At first this may seem incongruent with the emphasis placed upon the amplitude spectrum throughout this chapter. However, the discrepancy is essentially due to two factors. First, we only altered the phase of one FD to create our scrambled stimuli (as opposed to all phases when creating scrambled images). This means that phase was largely constant across recognisable and scrambled stimuli. Second, due to their complicated nature, images will have to be broken into many different frequencies with various interactions across phase in two dimensions. In contrast, for our relatively simple outlines it takes markedly few FDs to approach the original outline. As such, this should bias spectrum importance away from phase interactions, towards the amplitude of those phases when defining features. Nevertheless, the importance of the phase spectrum is easily demonstrated by simply observing the profound effect adjusting the phase of just one FD has on our shape stimuli. Therefore, we do not think the ideas put forward in this chapter and those presented earlier are inconsistent.

To conclude, we believe this study largely supports the choice of amplitude manipulations (as opposed to ‘number of FD’ manipulations) in the previous chapter. It also serves to legitimise our scrambled stimuli as adequate controls for their recognisable counterparts. We also note that these results indicate more generally that behavioural studies may be a useful preliminary exercise for determining suitable stimulus sets. For example, if we can behaviourally estimate the spacings between stimuli that participants perceive as linear increments, then those spacings could be used to parametrically vary stimuli for fMRI experiments. Such behavioural experiments could also be useful on their own merits, to test whether a given stimulus metric has any perceptual salience.

## **Chapter 7. Exploring the Role of Curvature for Neural Representations within Shape-Selective Lateral Occipital Cortex**

### **7.1 Abstract**

We conducted two fMRI experiments to probe neural shape representations primarily within Lateral Occipital brain regions of interest (ROIs) LO-1, LO-2 and LO. The first used radial frequency patterns to test whether amplitude (lobe size) or frequency (number of lobes) was most influential. We found only LO-1 showed hints of a frequency representation, whereas amplitude was well represented across all Lateral Occipital ROIs. However, amplitude, alongside additional stimulus similarity metrics, could be broken down into two discrete factors; ‘Lobe prominence’ and ‘Lobe Curvature’. These factors were both influential and implied Lateral Occipital shape representations may in part be based upon protrusions; both the degree to which a feature does protrude, and the nature of the protrusion (e.g. its breadth). The second study compared activation in the same ROIs to stimuli defined with points of maximal curvature change, the converse of those points (‘Not Curves’) and various other control conditions. This study tested the hypothesis that shape representations may be based upon points of curvature change. Intriguingly, whilst a general preference for coherent form emerged in Lateral Occipital ROIs (as would be expected), that preference tended towards the ‘Not Curves’ stimuli. This was contrary to our predictions, and so alternative explanations were explored. We found the ‘Not Curves’ shapes contained significantly more radial information than other conditions, therefore this again may point to a protrusion based representation in Lateral Occipital Cortex.

## 7.2 Introduction

In chapter 5 (Vernon et al., 2016), we explored shape representations across visual cortex and two dominant influences emerged. The first influence, which we termed the ‘Shape-profile’ factor, was highly predictive of responses in early visual cortex (namely V1-V3 and to a lesser extent V4). This was relatively straightforward to interpret as it captured the degree of spatial overlap between the shape stimuli used for those experiments. As such, that result is in accordance with the known retinotopic representations in those areas; shapes that cover similar portions of the visual field activate similar portions of early visual cortex.

Of greater interest was the second influence, which we termed the ‘Shape-complexity’ factor. This captured variance in more anterior extrastriate regions, namely the Lateral Occipital Complex (LOC); a region known to be selective for objects or shapes over stimuli without coherent form (Malach et al., 1995; Grill-Spector et al., 2001). Critically, this representation appeared to emerge within retinotopic areas LO-1 and LO-2 (Larsson and Heeger, 2006). These two areas contain full hemifield maps and lie between the dorsal component of V3 (V3d) and the more posterior/dorsal parts of the LOC (namely LO), with slight overlap between LO-1, LO-2 and LO sometimes observed (Sayres and Grill-Spector, 2008; Silson et al., 2016; Vernon et al., 2016). Within LO-1 we found a mixed representation; both Shape-profile and Shape-complexity factors could predict activity here to varying extents. In contrast, the representation within LO-2 appeared to be solely influenced by our Shape-complexity factor, in line with the LOC.

There are two key implications of these results. First, they demonstrate that regions such as LO-1 or LO-2, with known retinotopic organisation, can nevertheless show tunings for more abstract visual features like ‘Shape-complexity’. Second, this blend of retinotopic organisation alongside more abstract tunings could place LO-1 and LO-2 as potential preliminary processing stages for later extrastriate regions, most plausibly LO. Based upon this, we are conducting two studies to further elucidate the nature of our Shape-complexity factor. Understanding this representation could shed some light on both the underlying organisational principles within LO, as well as the roles that LO-1 and LO-2 play in forming that organisation.

An increase in Shape-complexity represents an increase in the amount of detail around a shape's contour, and this in turn will cause concomitant increases in 'curvature complexity'. That is, there will be more *frequent* curves (i.e. more troughs/ridges) and those curves will have greater *amplitude* (i.e. they will be more pronounced). Our previous studies could not disambiguate the influences of frequency and amplitude, and so this will be the first aim of our follow-up investigations. By using radial frequency (RF) patterns (Wilkinson et al., 1998), we can simultaneously and independently manipulate frequency information (number of lobes) and amplitude information (size of those lobes). This RF pattern experiment will use an identical paradigm to that utilised in earlier work (Vernon et al., 2016), allowing us to compare neural similarity in various visual areas with amplitude and frequency predictors. There is already evidence that amplitude manipulations are salient for LO (Drucker and Aguirre, 2009), and so we hope to replicate this finding. We also aim to extend it by searching for tunings either to frequency, or perhaps tunings to changes caused by the two influences combined.

We note that whilst conducting our experiment a similar paper was published (Salmela et al., 2016). RF patterns were manipulated in terms of amplitude and frequency, and then a representational similarity analysis was conducted across visual cortex. They argued for frequency tuning in areas V2d, V3d, V3AB and IPS0. However, there are a number of key methodological differences between their study and our own. Most critically, their regions of interest (ROIs) were defined using a probability atlas rather than on individual data as we typically do. Whilst this may suffice for earlier, larger ROIs, later regions such as LO-1 and LO-2 are small and (in our experience) their precise location can be relatively variable across participants. Furthermore, given that there are known dissociations between LO-1 and LO-2 (Silson et al., 2013), a finely demarcated boundary is essential. This has additional relevance as they later collapsed LO-1 and LO-2 into a single ROI, whereas we are interested in teasing these two regions apart. Their ROI boundaries may be further blurred as searchlight analysis was conducted, rather than comparing voxels within each ROI independently. Whilst popular, searchlight analysis does have the potential to both classify uninformative voxels as informative, and miss patterns of weakly informative voxels (Etzel et al., 2013). As such, we prefer the approach (where possible) of defining voxels based upon independent data to ensure

that we do not miss or falsely classify voxels at least at the level of the ROI. Finally, we note that low level influences of spatial frequency and contrast energy could account for a large proportion of their explained variance. In our study we aim to control for stimulus size (perimeter, area), which should serve to model out these low-level confounds. In sum, for the reasons discussed above we believe that our approach should yield results that will still be a valuable addition to the literature.

Our second follow-up investigation will take a broader approach, assessing tunings for curvature more generally. Specifically, we will render shapes using either the full outline, points of peak curvature only ('curves'), the converse of those points ('not curves'; removing curvature whilst maintaining shape information), rotated curvature points (maintaining curvature whilst removing shape information) or finally completely scrambled outlines (removing curvature and shape information). These shapes will be presented in a block-design fMRI experiment, allowing us to compare responses to each condition. We generally expect 'shape information' conditions to elicit greater activity than 'no shape information' conditions in Lateral Occipital areas, based upon the known shape sensitivity in this region. However, if the more abstract/Shape-complexity tunings observed in LO-2/LO do in fact reflect responses to curvature content, then we would expect an additional hierarchy of activation such that curvature conditions will elicit greater activity than minimal curvature conditions. Based upon this, we expect 'full outlines' to elicit greater activation than 'curves' as they contain the most explicit shape information. 'Curves' should then elicit greater activation than 'not curves' and (likely to a greater extent) 'rotated curves' due to the lack of curvature content and shape content in those two conditions respectively. Finally, 'curves' should elicit considerably more activation than 'scrambled' shapes as they lack both curvature and shape information.

## **7.3 General Methods**

### **7.3.1 Participants**

Both studies used the same pool of twelve participants (Mean age 25.75, SD 5.05; 6 Males; all gave informed consent), recruited from the University of York Psychology department. Each participant underwent one high-resolution structural scanning session, one retinotopic-mapping session, one LOC localiser session and two main functional sessions, totalling 5.25 hours scanning per participant.

### 7.3.2 Preliminary Data Acquisition and Analysis

Structural, retinotopic mapping and LOC functional localiser scans were conducted as described previously (Chapter 2, Methods, sections 2.3 to 2.6). Regions V1-V4, LO-1 and LO-2 were retinotopically identified in each participant.

To define LO and pFs we used a slightly different method to that utilised in the previous chapters. Specifically, we took the cluster-corrected ( $Z > 5.0$ ,  $p < .001$ ) significant activity from the LOC localiser and rendered it on each individual's surface, restricted to grey matter only. In 11 of the 12 participants, we identified two discrete bilateral clusters in Lateral Occipital cortex, which likely correspond to LOa and LOb (Vinberg and Grill-Spector, 2008); in the 12<sup>th</sup> participant only LOb was identified. However, activation in LOa was less robust than LOb and it typically showed considerable or complete overlap with LO-2, meaning most voxels would be redundant (analysed twice). As such, the larger cluster corresponding to LOb was taken as our LO ROI. A third more anterior/ventral cluster was identified in all participants, corresponding to the Posterior Fusiform Gyrus (pFs). This approach has similar advantages to the 'sphere-on-peak-voxel' approach used previously (Vernon et al., 2016), namely it minimises LO-2 and LO overlap. However, it also takes into account individual variation in terms of the size of LO, meaning that our selection of voxel choices should be more accurate.

All ROIs were collapsed (concatenated) across hemisphere as we had no *a priori* reasons to look for hemispheric differences.

### 7.3.3 ROI Restriction

Due to the rapidly increasing receptive field sizes across visual cortex (Dumoulin and Wandell, 2008; Amano et al., 2009), earlier visual areas will have far fewer neurons stimulated than later regions. Due to this, we aimed to restrict our ROIs to the stimulus representation using an additional stimulus localiser scan.

The stimulus localiser (TR = 3000ms, TE = 30ms, voxel size =  $2 \times 2 \times 2 \text{mm}^3$ , flip angle =  $90^\circ$ , matrix size  $96 \times 96 \times 38$ , FOV = 19.2cm) used an ABAB block design (block length 15sec) in one 4min run, with 9 seconds prepended at the start to ensure magnetisation reached a steady state. The blocks alternated between annulus-shaped texture patterns (inner radius  $0.5^\circ$ , outer radius  $5^\circ$ ) and the corresponding converse patterns (full screen textures with annulus masked). The outer radius size was chosen

to ensure that its area would encompass all possible shapes across both studies. We excluded the inner  $1^\circ$  of visual angle as this region was typically not stimulated by our stimuli, and the foveal region represents a large proportion of visual cortex. Masked regions (e.g. centre and surround of annulus) were set to mid-grey.

To create our texture patterns, we used Voronoi diagrams generated from random dot patterns; the number of dots was chosen such that there would be one dot every  $1^\circ$  of visual angle<sup>2</sup> on average. A Voronoi diagram splits the pattern into cells, such that each cell contains the region of all points lying closest to a given dot. For each Voronoi diagram, we created four versions in which the cell colours were randomly cycled between linearly spaced greyscale values (RGB triplet values: 0 (black), 85, 170, 255 (white)). During a given block the texture cycled between all four versions of a given Voronoi diagram (updating every 0.25sec), before switching to a new diagram (a given diagram was never repeated after its first presentation sequence). We reasoned that these diagrams should robustly stimulate early visual cortex, as they contain large amounts of contrast and edge information at random spatial positions and orientations. In addition, the diagrams contain shape information (due to the random nature of the cells) to stimulate Lateral Occipital cortex. Participants passively maintained fixation on a red cross ( $0.5\text{deg}$ . visual angle) during the run.

Data were analysed using FEAT (Worsley, 2001). The first three volumes of the run were discarded, the high-pass filter cut-off point was set to 60sec (correcting for low frequency drift), FILM prewhitening and spatial smoothing (Gaussian kernel with FWHM of 4mm; twice voxel size) were used, motion was corrected for (motion parameters were also entered as confound covariates). The annulus condition was modelled (convolved with double-gamma HRF) against the surround as baseline. To identify significant activation, we used fixed effects analysis with cluster correction ( $Z > 2.3$ ,  $p < .05$ ), all data were retained in the high-resolution structural space.

Once significant activation was identified we aimed to simply restrict ROIs based upon that activation. However, unfortunately in most participants the activation showed a striking cut-off past V3, sometimes slightly extending to V4 and LO-1, but not to great extents. We believe that receptive field sizes in the inactive regions must be large enough to be stimulated by both the annulus and surround texture, such that when contrasted against each other no voxels emerge as significant. Due to this, only

V1-V3 could be restricted based upon this stimulus localiser (and consequently they have been for all subsequent analyses). Nevertheless, the very fact that receptive field sizes in later regions are too large to be selectively stimulated by our localiser implies that few voxels would have been restricted anyway, if a different annulus size had been selected. As such, we do not believe that this is a major confound in our analysis.

## 7.4 Methods - Radial Frequency Experiment (Expt. 1)

### 7.4.1 Stimulus Creation

The first experiment used radial frequency (RF) patterns, defined using the formula:

$$r_0(1 + A(\sin(\omega\theta + \phi)))$$

Here, theta ( $\theta$ ) represents the angles around a circle's perimeter, allowing the sinusoidal modulation of that perimeter by altering frequency ( $\omega$ ) and amplitude ( $A$ ), rotation can be set by altering phase ( $\phi$ ). The mean radius ( $r_0$ ) governs the average size of the stimulus, and was set to 2.5 degrees of visual angle. To display the shapes, the outer contours were rendered against a mid-grey background using the radial fourth derivative of a Gaussian (Wilkinson et al., 1998) at 50% contrast (as outlined in Chapter 2, Methods, section 2.7), yielding a peak spatial frequency of 3.31 cycles per degree. Reverse-polarity versions were also created by multiplying the Gaussian function by -1 (inverting it).

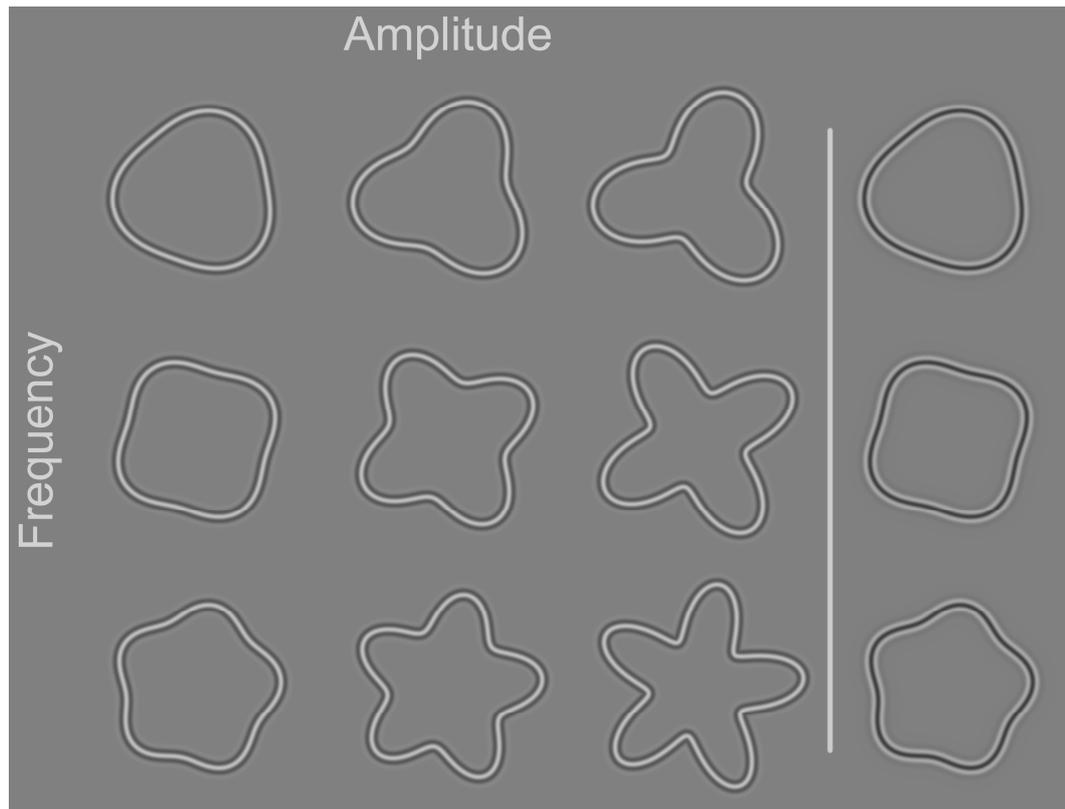
To create a 3x3 (frequency by amplitude) stimulus set, we set frequency to 3, 4 or 5 lobes. The RF patterns were first oriented such that they either had one lobe pointing vertically upwards (RF3, RF5) or one indentation pointing vertically upwards (RF4), then additional rotation was applied to display the RF patterns at 4 discrete phases (60°, 120°, 240°, 300°), all of which avoided a vertical or horizontal axis of symmetry in case that may be atypically salient.

The appropriate range of amplitudes to use was more difficult to determine, as there are no discrete intervals as is the case with frequency. To address this, we used unconstrained nonlinear optimisation (Mathworks Matlab R2012a function *fminsearch*) to match amplitude changes to frequency changes in terms of low-level visual details. Specifically, for a given pair of amplitudes (i.e. the lower and upper

amplitudes; the middle amplitude is then generated from the average), we generated the 3x3 RF pattern stimulus set at all four phases/orientations (i.e. 36 stimuli total). GIST descriptors (Oliva and Torralba, 2001) were then computed for each unique stimulus; these convolve an image with various Gabor filters, providing metrics detailing the amount of information present at various orientations and spatial scales (akin to ‘low-level’ visual information). We chose GIST descriptors as we previously found them to be the best metric for assessing shape similarity in lower-level visual areas (namely V1-V3) (Vernon et al., 2016).

For the set of stimuli at a given phase value, we computed the GIST similarity between all (36) pairwise combinations of shapes. In a given stimulus pair (Shape1-Shape2), this is simply the average absolute difference between the GIST descriptors for Shape1 and the GIST descriptors for Shape2. We also calculated analogous frequency and amplitude similarity metrics (e.g. if Shape1 has frequency 3, and Shape2 has frequency 5, frequency similarity for that pair will be  $|3-5| = 2$ ). We then ran partial correlations correlating frequency and GIST similarities (controlling amplitude), plus amplitude and GIST similarities (controlling frequency). When this process had been repeated across all four phase values, we averaged the (Fisher Z transformed) correlations across phase to generalise similarity across all orientations (as orientation is incidental to our hypotheses). Our nonlinear optimisation aimed to minimise the absolute difference between the average ‘frequency-GIST’ and ‘amplitude-GIST’ correlations, by tweaking the lower and upper amplitudes.

When run, this produced amplitudes of 0.09, 0.23 and 0.36 with frequency-GIST and amplitude-GIST partial correlations of  $r = .5405$  and  $r = .5406$  respectively. The final stimuli can be seen in Figure 7.1. Essentially this should mean that moving along the amplitude dimension elicits roughly equivalent low-level visual changes as moving along the frequency dimension. This is critically important for our design, as we aim to compare the importance of the amplitude and frequency dimensions across visual cortex. If, for example, the amplitude changes we had chosen had been more salient for early visual cortex (e.g. due to orientation or contrast energy), then such differences could have produced confounding influences across the visual hierarchy. By constraining the dimensions in a low-level sense, we should allow more subtle representations to emerge in our extrastriate regions of interest.



**Figure 7.1.** Radial Frequency (RF) Stimuli. The nine RF stimuli used in this experiment are depicted (at a sample orientation) in the panel to the left. These stimuli are split into orthogonal dimensions of amplitude (size of a lobe) and frequency (number of lobes). The panel to the right depicts example oddball (reversed polarity) stimuli, which participants were instructed to respond to.

#### 7.4.2 Data Acquisition and Analysis

The RF stimuli were presented centrally in a rapid event-related fMRI design ( $TR = 2000\text{ms}$ ,  $TE = 30\text{ms}$ , voxel size =  $2 \times 2 \times 2.5\text{mm}^3$ , flip angle =  $77^\circ$ , matrix size  $96 \times 96 \times 26$ , FOV =  $19.2\text{cm}$ ), with eight 5min runs. An additional 10 seconds were prepended to each run to ensure magnetisation reached a steady state, plus 20 seconds were appended to ensure we captured the full hemodynamic response for the final stimuli presented (leading to a total run length of 330sec).

Within a given run, each stimulus was presented eight times (totalling 72 stimuli per run) with a stimulus presentation time of 0.6sec. The stimuli were presented at each of the four phase values (orientations) twice in random order; this was to minimise the influences of low-level visual properties on shape processing (as otherwise e.g. all RF3 patterns would have lines/edges at similar orientations, confounding more general shape similarity). Stimulus presentation order was counter-balanced and

optimised (jittered) using Optseq2 (<https://surfer.nmr.mgh.harvard.edu/optseq>); eight stimulus presentation sequences were generated (one per run) and used in order of most-to-least efficient (ensuring most efficient runs were used when participants were most alert). The median inter-stimulus interval was 3.4sec (interquartile range 2.4-5.4sec, range 1.4-24.4sec).

Participants maintained fixation on a red cross (0.25deg. visual angle) and performed an oddball task; they were instructed to press a button every time they saw an RF stimulus with reversed polarity as shown in Figure 7.1 (participants were familiarised with ‘normal’ and reversed RF patterns beforehand). In each run 12 of the 72 stimulus presentations were oddballs. The 72 presentations were split into 6 sets of 12, in each set there could either be 0, 1 or 2 oddballs, allocated randomly. This ensured that oddballs would be spread relatively evenly throughout the run length. In addition, we ensured that the very first stimulus presentation could not be an oddball, and there would be no sequentially repeated oddballs.

Data were analysed using FEAT (Worsley, 2001). The first five volumes of each run were discarded, the high-pass filter cut-off point was set to 100sec (correcting for low frequency drift), FILM prewhitening and spatial smoothing (Gaussian kernel with FWHM of 4mm; twice voxel size) were used, motion was corrected for (motion parameters were also entered as confound covariates). All nine stimuli were entered as separate explanatory variables (EVs), convolved with a double-gamma HRF, and contrasts were run to compare each individual stimulus to baseline. We also included oddball events as an additional confound EV, to control for any variance they might account for (e.g. sudden attention or participant movement when responding). Runs were combined within-participant using fixed-effects analysis with cluster correction ( $Z > 2.3$ ,  $p < .05$ ), all data were retained in the high-resolution structural space.

### **7.4.3 Correlation Analysis and Shape Similarity Metrics**

As in chapter 5, our analysis aimed to explore neural representations by comparing neural similarity to various shape similarity metrics (predictors) that are based upon known physical properties of our shape stimuli. Briefly, the neural similarity matrix for a given ROI is the set of correlations generated by correlating patterns of activity within that ROI across all (36) pairwise shape combinations (i.e. correlate activity pattern elicited in V1 by Shape1 with that elicited by Shape2, etc.). If a neural

similarity matrix for a given region correlates with a given predictor variable, then it implies that the predictor is capturing some aspect of the shape representation within that region. To create a predictor variable for a given metric (e.g. frequency), we simply take the set of absolute differences between the metrics generated across all pairwise combinations of shapes, then subtract it from 0 so larger numbers represent greater similarity (e.g. if Shape1 has frequency 3 and Shape2 has frequency 4, the matrix element for that combination will be  $0 - |3 - 4| = -1$ ).

The first predictor variables to include are RF amplitude and frequency, as they formed the hypotheses for this study. We also required predictor variables to control for stimulus size, as manipulations in amplitude are unavoidably confounded by the spatial extent the stimuli take up (larger amplitudes result in shapes with larger maximum diameters). There will also be general increases in other low-level visual properties (e.g. contrast energy) as a function of stimulus size. As such, predictors for both stimulus area and perimeter were included. Finally, we wanted to include additional predictors that might capture variance not covered by amplitude and frequency alone (or are related to interactions between amplitude and frequency). Naturally, we first assessed the Shape-complexity predictors used in our previous study (Vernon et al., 2016), however both ‘Number of Fourier Descriptors’ and ‘Minima/Maxima’ would be identical to frequency in the current stimulus set. Furthermore, over 90% of the variance in each of the original ‘Compactness’ and ‘Convex Perimeter’ predictors could be captured by stimulus size (area and perimeter) alone. This implies that our original predictors are not as appropriate for RF stimuli and novel predictors are warranted. Accordingly, four new predictors were created:

1. *Lobe Angle*: the angle formed between lines connecting a given lobe peak with its two nearest indentations. This angle should decrease both as a function of frequency and amplitude

2. *Radius of Curvature*: this is the radius of the circle whose curvature matches the curvature of an RF stimulus at the peak of a lobe. An RF stimulus with ‘broader’ lobes should elicit a larger radius of curvature. This was calculated in polar coordinates with the following formula:

$$R = \frac{(r^2 + r_{\theta}^2)^{3/2}}{|r^2 + 2r_{\theta}^2 - rr_{\theta\theta}|} \quad \left( \text{where } r_{\theta} = dr/d\theta, r_{\theta\theta} = d^2r/d\theta^2 \right)$$

3. *Contour Smoothness*: this was calculated by sampling an RF stimulus every  $0.5^{\circ}$  visual angle around the contour, then calculating the average absolute angle between all pairs of points. A smoother pattern should have fewer large deviations in angle and so the resultant average should be lower.
4. *Concentricity*: this was the average absolute angle between the tangent at a given point on the RF stimulus and the radial line (line passing through pole). A circle would be perfectly concentric and so all angles should be  $90^{\circ}$ , as frequency and amplitude increase an RF pattern becomes less concentric and more radial, resulting in smaller average angles. This was calculated in polar coordinates with the following formula:

$$\Psi = \left| \tan^{-1} \left( \frac{r}{dr/d\theta} \right) \right|$$

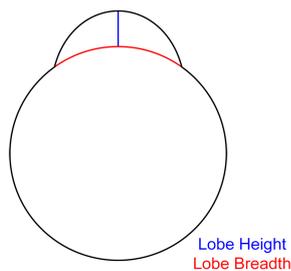
As we wanted to control for stimulus size, we regressed both stimulus area and perimeter predictors on all additional predictors above, saving out the standardised residuals. All references to the additional predictors (including amplitude and frequency) will henceforth refer instead to the predictor’s residuals, as these should only contain variance that is not accounted for by stimulus size. Note that the two stimulus size predictors did account for a large amount of the variance in amplitude (91.6%), but it was retained due to its importance for our hypotheses. The variance accounted for was lower in other predictors (Frequency: 5.7%, Lobe Angle: 78.0%, Radius of Curvature: 26.5%, Contour Smoothness: 23.5%, Concentricity: 78.3%).

For analysis, all correlations were transformed to Fisher Z-Scores before averaging or statistical testing.

#### 7.4.4 Visualising Predictors

As a means to actually *visualise* how these new predictors might ‘describe’ RF patterns, we performed some computational modelling on a sample lobe.

Specifically, we created a circle whose radius matched that of our RF patterns, and rendered a single lobe upon its contour (Figure 7.2). The lobe was defined in terms of height (varying from 1/15 to 1/3 of the circle’s radius) and breadth (varying from 1/15 to 1/3 of the circle’s circumference), by manipulating the amplitude (from 1/15 to 1/3) and frequency (from 1.5 to 7.5) of a sine wave for height and breadth respectively. Crucially, this method allows us to freely vary the properties of the lobe, without the constraints inherent in RF patterns (for example, we are not limited to integer frequencies). To create a shape space, we iterated over 1,000 linearly spaced height intervals, and at each height interval we iterated over 1,000 linearly spaced breadth intervals (creating 1,000,000 stimuli total).

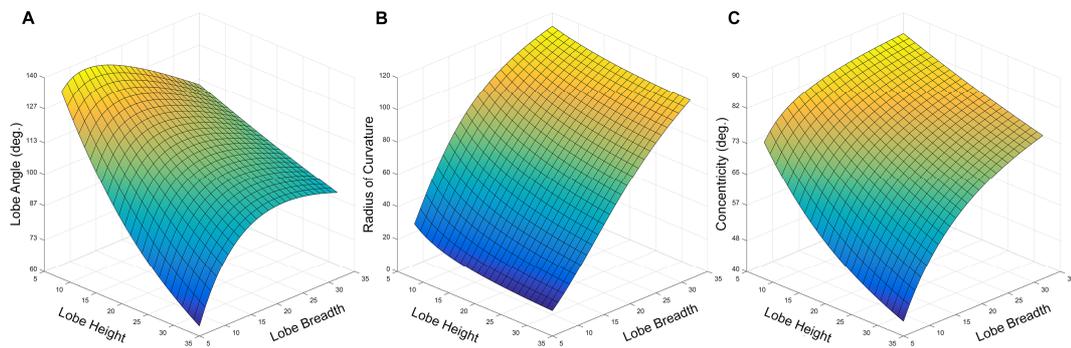


**Figure 7.2.** Sample lobe. To explore our predictor variables, we modelled them based upon a circular stimulus rendered with a single lobe. This lobe can independently vary in height and in breadth, between 1/15 and 1/3 of the circle’s radius and circumference respectively.

For each stimulus in this shape space, we then calculated lobe angle and radius of curvature as described above. We also calculated concentricity, however only based upon the points within the lobe itself. Unfortunately, we could not calculate contour smoothness as that predictor is critically dependent upon the entire contour, not just a single lobe. We can then plot each predictor as a function of lobe height and lobe breadth (Figure 7.3).

Generally, we see differing linear and non-linear relationships between the predictors and lobe properties. The linear relationships can at least be quantified by Pearson’s correlation coefficient ( $r$ ). First, for lobe angle (Figure 7.3A), we see a strong, almost linear relationship with lobe height ( $r = -.86$ ), however this is moderated by a non-linear relationship with lobe breadth ( $r = .07$ ). Conversely, radius of curvature (Figure 7.3B) is almost entirely governed by a linear relationship with lobe breadth ( $r = .96$ ), with only minor influences from lobe height ( $r = -.23$ ). Finally,

concentricity (Figure 7.3C) appears to be broadly linearly related to both dimensions (lobe height:  $r = -.61$ ; lobe breadth:  $r = .73$ ).



**Figure 7.3.** RF predictor variables. Here we plot lobe angle (A), radius of curvature (B) and concentricity (C) against lobe height and lobe breadth. Height and Breadth are expressed as a percentage of the circle’s radius and circumference respectively.

These results highlight two key points. First, they demonstrate that even though the variables are all fundamentally based upon the same shape, they can vary considerably, both in linear and non-linear fashions. However, the fact that non-linear relationships exist highlights both a strength and a limitation of the general approach taken so far in this thesis, and in this chapter.

Neuronal responses will have highly non-linear responses to stimuli, yet simple correlations might fail to identify such relationships because they assume linearity. Indeed, there is evidently some relationship between lobe breadth and lobe angle (Figure 7.3A), but correlations fail to capture it. Critically however, the predictor itself can act as an intermediary in this situation. If both a predictor and the neural activity change in the *same* non-linear fashion, then the predictor can in a sense ‘absorb’ that non-linearity, allowing relationships to reveal themselves through more conventional statistics. This makes the predictor approach valuable in probing neural representations across visual cortex.

However, we have previously used principal component analysis (PCA) to reduce the predictors into a common dimensionality, and will do so again later in this chapter. PCA is fundamentally based upon correlations between its input, and so it attempts to identify *linear* relationships between the predictors. This means that ultimately PCA may fail to detect the more nuanced relationships across stimuli that likely exist. As such, a goal for future work should be to develop analysis strategies

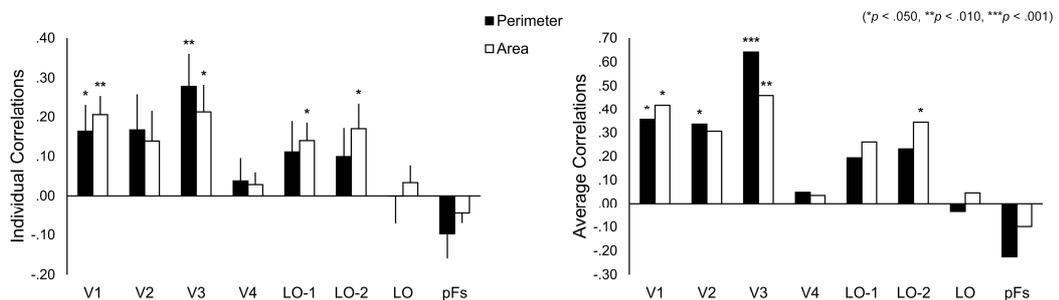
that let these non-linear relationships emerge, as doing so may allow us to better capture the dimensionality of ‘shape space’ in Lateral Occipital cortex.

## 7.5 Results - Radial Frequency Experiment (Expt. 1)

### 7.5.1 Stimulus Size (Figure 7.4)

As an initial step we assessed the influence of stimulus size, which we have chosen to model out of all future analyses. Two key approaches will be used to evaluate each predictor. First, as a more conservative option we will correlate the predictors with the neural similarity matrices of each individual in each ROI, testing whether the resultant average correlations differ significantly from zero (via one-sample t-tests). Second, we will average neural similarity matrices across participants within each ROI and look for significant correlations between these average neural similarity matrices and the predictors. In previous work (Vernon et al., 2016), we found that averaging neural similarity across participants reduced individual variability as a source of noise, providing a cleaner picture of the results.

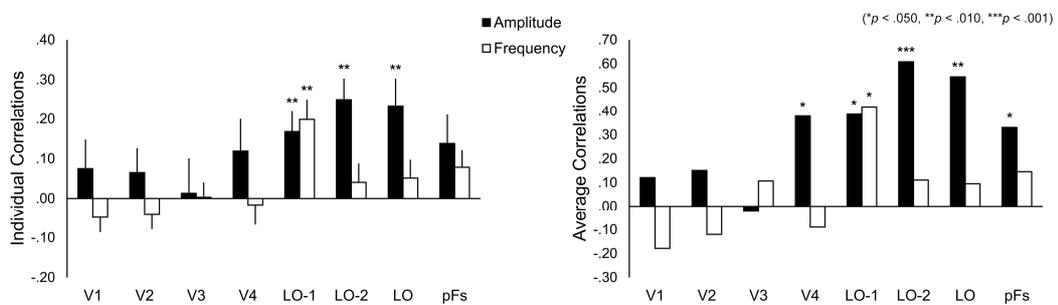
Across individuals, we found that stimulus perimeter only correlated with V1 ( $t(11) = 2.38, p = .037$ ) and V3 ( $t(11) = 3.51, p = .005$ ). Stimulus area correlated with V1 ( $t(11) = 4.23, p = .001$ ), V3 ( $t(11) = 3.06, p = .011$ ), LO-1 ( $t(11) = 3.09, p = .010$ ) and LO-2 ( $t(11) = 2.61, p = .024$ ). However, when looking at average neural similarity only correlations in V1-V3 were significant for perimeter (V1:  $r = .36, p = .033$ ; V2:  $r = .34, p = .045$ ; V3:  $r = .64, p = 2.5 \times 10^{-5}$ ), and V1, V3 and LO-2 for area (V1:  $r = .42, p = .011$ ; V3:  $r = .46, p = .005$ ; LO-2:  $r = .34, p = .040$ ).



**Figure 7.4.** Correlations between RF pattern perimeter and RF pattern area with neural similarity. Left panel: average correlations with individual participant’s neural similarity matrices, error bars represent the SEM. Right panel: correlations with the average neural similarity matrix for each ROI.

### 7.5.2 Amplitude and Frequency (Figure 7.5)

We now turn to the main predictors of interest - amplitude and frequency. To reiterate, these (and all subsequent) predictors have stimulus area and perimeter regressed out. Across individuals, we find amplitude only correlates with Lateral Occipital ROIs LO-1 ( $t(11) = 3.31, p = .007$ ), LO-2 ( $t(11) = 4.79, p = .001$ ) and LO ( $t(11) = 3.46, p = .005$ ). Intriguingly, the only ROI to show a significant correlation with frequency is LO-1 ( $t(11) = 3.89, p = .003$ ). The same is true when looking at average neural similarity; only LO-1/frequency is significant ( $r = .42, p = .011$ ). For amplitude, we again see significant correlations in LO-1 ( $r = .39, p = .020$ ), LO-2 ( $r = .61, p = 8.3 \times 10^{-5}$ ) and LO ( $r = .55, p = .001$ ), but now also V4 ( $r = .38, p = .023$ ) and pFs ( $r = .33, p = .049$ ).



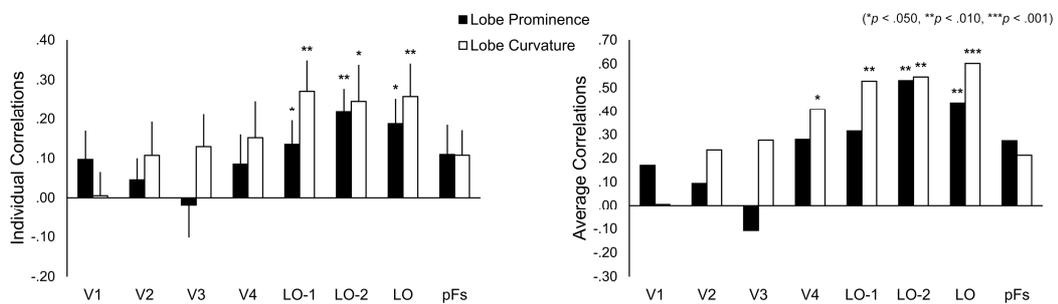
**Figure 7.5.** Correlations between RF pattern amplitude and RF pattern frequency with neural similarity. Left panel: average correlations with individual participant's neural similarity matrices, error bars represent the SEM. Right panel: correlations with the average neural similarity matrix for each ROI.

### 7.5.3 Additional influences (Figure 7.6)

Finally, we explored the four novel predictors (lobe angle, radius of curvature, contour smoothness and concentricity) to see if they can capture additional variance. However, as these predictors and amplitude/frequency are all intended to assess similar properties, we first ran them through principal components analysis (PCA) to reduce the dimensionality of the predictor set. Preliminary checks showed that frequency had relatively low correlations with the other predictors, as well as a low communality after extraction (.405 versus .710 for next lowest communality), implying that it was a relatively independent predictor. As such, frequency was excluded from the PCA. The PCA with five predictors (amplitude plus the four described above) now produced a clear two factor solution accounting for 89.52% of the variance (59.40% and 30.12% for the two factors respectively). Both factors had

eigenvalues greater than 1 in accordance with Kaiser's (1960) criterion, the scree plot showed a point of inflection after the second factor and Horn's (1965) Parallel Analysis also suggested a two factor solution. The factors were extracted with Varimax (orthogonal) rotation (we first attempted Direct Oblimin rotation allowing correlated factors, but resultant correlations between factors was relatively low;  $r = .23$ ), and they were saved as Anderson-Rubin variables to ensure factor independence.

Amplitude and lobe angle loaded highly on factor 1 (.967, .969 respectively), whereas radius of curvature and contour smoothness loaded highly on factor 2 (.891, .899 respectively). Concentricity showed moderate loadings on both factors (.765, .600 respectively). All other loadings were minimal (all  $< .182$ ). From this, it seems plausible that factor 1 captures the degree to which lobes 'protrude' from the base shape, so we will term it 'Lobe prominence'. Factor 2 instead seems to be capturing details about the nature of the lobe, perhaps smoother more rounded curves versus abrupt sharper turns, so we will term it 'Lobe curvature'. We can now explore these two factors with respect to the neural similarity matrices.



**Figure 7.6.** Correlations between the 'Lobe prominence' and 'Lobe curvature' factors with neural similarity. Left panel: average correlations with individual participant's neural similarity matrices, error bars represent the SEM. Right panel: correlations with the average neural similarity matrix for each ROI.

Across individuals, we see a hint of a 'Lobe prominence' influence in LO-1 ( $t(11) = 2.22, p = .049$ ), but it appears to play a greater role for LO-2 ( $t(11) = 3.70, p = .004$ ) and LO ( $t(11) = 2.91, p = .014$ ). However, 'Lobe curvature' is broadly equally represented across LO-1 ( $t(11) = 3.58, p = .004$ ), LO-2 ( $t(11) = 2.73, p = .020$ ) and LO ( $t(11) = 3.21, p = .008$ ). Results for the average neural similarity matrices are similar. 'Lobe prominence' correlates with LO-2 ( $r = .53, p = .001$ ) and

LO ( $r = .44, p = .008$ ). ‘Lobe curvature’ correlates with LO-1 ( $r = .53, p = .001$ ), LO-2 ( $r = .55, p = .001$ ), LO ( $r = .60, p = 1.0 \times 10^{-4}$ ) and now also V4 ( $r = .41, p = .013$ ).

#### 7.5.4 Extending the ventral stream; VO-1/2

As in earlier work (Vernon et al., 2016), we did also test additional ventral (retinotopically defined) ROIs VO-1/2, collapsed into a single VO ROI. This is simply to test whether any of the above results may be accounted for by the increased receptive field sizes of more anterior visual areas. However, none of the predictors correlated with neural similarity in VO, whether at the individual (all  $p > .109$ ) or average (all  $p > .112$ ) level. This implies the results discussed above do indeed reflect the tuning characteristics of the respective regions, and are not a general consequence of e.g. increasing receptive field sizes.

### 7.6 Discussion - Radial Frequency Experiment (Expt. 1)

To briefly summarise our results, we generally found that stimulus size was most influential for lower level visual areas (namely V1 and V3), with some influence on later LO-1 and LO-2 ROIs. Amplitude was key for all Lateral Occipital ROIs (and perhaps V4), whereas frequency only appeared to influence LO-1. Finally, we were able to generate two factors reflecting different (independent) properties of our RF stimuli, termed ‘Lobe prominence’ and ‘Lobe curvature’. Both appeared to play a role in (and were largely specific to) Lateral Occipital neural representations.

The stimulus size result is relatively unsurprising, given that the regions in which size correlates with are all known retinotopic areas and therefore spatial extent should influence which neurons are stimulated. We note however that these results do not really present a clean transition between retinotopic and more abstract predictors as found previously (Vernon et al., 2016). There are two potential reasons for this. First, in this study we aimed to control for low-level influences by presenting stimuli at four different orientations, this would ‘blur’ retinotopic representations leaving only a coarse trace linked to stimulus size. Second, the stimuli used in this study have a far more uniform profile than previous stimuli, all being broadly circular, and so there will be less ‘spatial variance’ with which we can model retinotopic differences.

The main aim of this study was to disambiguate the influences of amplitude and frequency on neural representations in Lateral Occipital cortex, asking which was key (if not both). The current results come out resoundingly in favour of amplitude. Indeed, the amplitude results strongly mirror the pattern identified for Shape-complexity in previous work (Vernon et al., 2016), suggesting the results could be analogous. However, this is qualified to some extent by the factor results, discussed shortly.

First, it was surprising that frequency only correlated with LO-1, and no other ROIs. This may in part explain the dissociation identified between LO-1 and LO-2 (Silson et al., 2013), as the results imply that LO-1 does play some independent role. In Silson et al., LO-1 was found to play a causal role in orientation discrimination, supported by previous findings showing second-order orientation sensitivity in this region (Larsson et al., 2006; Montaser-Kouhsari et al., 2007). It is plausible that some more ‘abstracted’ orientation processing underlies our results. As frequency increases, there will be more orientation energy around the shape’s contour. This did not correlate with lower level visual areas, likely because the shapes were presented at different orientations so systematic links between orientation and spatial location would be blurred. As such, it is possible that rather than strictly retinotopic orientation processing, LO-1 instead processes orientations anchored to specific objects. That is, LO-1 might process the greater orientation energy of e.g. an RF4 pattern over an RF3 pattern regardless of its current rotation. Such a feature would prove useful as a preliminary shape processing stage, if that is LO-1’s role, it would also serve to disambiguate LO-1 from the known orientation processing in earlier visual areas (Hubel and Wiesel, 1959).

Turning to our factor results, the fact that we found two independent influences makes the amplitude result less straightforward to interpret. The ‘Lobe prominence’ dimension appeared to capture the degree to which lobes protrude, and this was only consistently influential for LO-2 and LO. Given that amplitude loads highly on this factor, this again reinforces the LO-1/LO-2 dissociation, as LO-2 (but not LO-1) was found to be causally implicated in an amplitude-related shape-processing task (Silson et al., 2013). It also corroborates previous findings of amplitude salience in LO (Drucker and Aguirre, 2009). However, this does not imply that ‘amplitude’ manipulations entirely underlie our previous results (Vernon et al., 2016), as ‘Lobe

curvature’ also proved to be as influential, if not more so. ‘Lobe curvature’ appeared to capture differences between smoother, more rounded lobes versus sharper, acute lobes. This factor correlated well with all Lateral Occipital ROIs (LO-1, LO-2 and LO), plus V4 to a limited extent. Combined, these influences place the representation within Lateral Occipital ROIs as one that is based on shape protrusions; both the prominence (or perhaps salience) of those protrusions and also the nature of them. Such a finding is in accordance with evidence that novel shapes with similar shape features (e.g. spikes vs. rounded lobes etc.) share similar representations in LO (Op de Beeck et al., 2008b), and it may also suggest an explanation for why convexities are preferentially responded to (both behaviourally and neurophysiologically) compared to concavities (Habak et al., 2004; Poirier and Wilson, 2007; Haushofer et al., 2008a; Kurki et al., 2014). It also reinforces our previous work showing shape complexity is influential for LO’s neural representations (Vernon et al., 2016).

The existence of two factors may also serve to explain why we previously found certain results differed between the ‘Standard’ (recognisable) and ‘Scrambled’ (unrecognisable) shape stimulus sets (Vernon et al., 2016). Most notably, V4 showed a strong retinotopic response profile for standard stimuli, but more mixed responses with scrambled stimuli. In the current study, only ‘Lobe curvature’ correlated with V4, and so it is possible that by ‘scrambling’ our previous stimuli we inadvertently biased the stimuli into making whatever features underpin the ‘Lobe curvature’ factor more influential than those that underpin ‘Lobe prominence’. Given that both factors correlate well with LO, such a bias would not have changed our core results. If ‘Lobe curvature’ is based upon rounder versus acute curves, then V4’s current result may be based in part on the concentricity of our stimuli (more rounded lobes should make the stimuli more concentric, less rounded lobes would make them more radial). Early work shows that V4 responds to both radial and concentric stimuli (Wilkinson et al., 2000), and so it makes some sense that V4’s response profile would be modulated specifically by ‘Lobe curvature’.

Finally, our results here differed considerably from those in a recent paper with a comparable design (Salmela et al., 2016). Most notably, we found frequency only correlated with LO-1 (which they also found to a limited extent), and that amplitude could capture significant amounts of variance in Lateral Occipital ROIs. Certain limitations of this paper were discussed earlier (e.g. not controlling for low-level

visual properties), and these could underlie the differences found. Alternatively, it is possible that the representations identified here are on a smaller resolution than Salmela et al. could identify. We used a considerably smaller voxel size (2x2x2.5 vs. their 3.1x3.1x3.0) and this may have given us the required resolution to identify the above results.

To conclude this section, we essentially identified four discrete stimulus representations, which had dissociable influences across visual cortex. First, we have stimulus size predominating in earlier visual areas, then there is a ‘frequency’ representation in LO-1 that could potentially represent abstracted orientation information. Next, across all Lateral Occipital ROIs (and to a lesser extent V4) there is a ‘Lobe curvature’ representation, potentially capturing some aspect of curvature smoothness or concentricity. Finally, in the more anterior Lateral Occipital ROIs (LO-2 and LO) there is a ‘Lobe prominence’ representation that appears to capture the degree to which lobes (or protrusions) actually protrude.

## **7.7 Methods – Gabor Element Experiment (Expt. 2)**

### **7.7.1 Stimulus Creation**

Our second experiment aimed to test for curvature tuning more generally. One hypothesis drawn from our previous work (Vernon et al., 2016) was that LO may respond to shapes primarily based upon points of peak curvature (minima/maxima). If so, then rendering a shape with just those points (as contour fragments) should elicit greater activity than if the converse of those points were rendered.

To create our stimulus set, we took 42 shapes from the Snodgrass & Vanderwart (1980) images that had been converted into silhouettes by DeWinter & Wagemans (2004). These shapes were resized to an area of 16° visual angle<sup>2</sup>, then converted into a set of Fourier Descriptors (FDs) (Zahn and Roskies, 1972) as described previously (Vernon et al., 2016). As we were aiming to render the shapes as contour fragments with just points of peak curvature showing, we needed to reduce the detail around the outer contour (otherwise the contours could be too detailed to split into curvature versus straighter segments). To do this, all shapes were constrained to just 6 discrete FDs, limiting contour complexity. We also wanted to ensure the shapes were completely unrecognisable as otherwise certain points of the shape may be unintentionally salient (for example we may recognise ears on a rabbit from curves

alone). This was addressed by ‘scrambling’ the shapes as in Vernon et al. (2016); specifically, we rotated the FD with greatest magnitude in each shape through  $90^\circ$  counter-clockwise.

We now have a stimulus set that contains completely unrecognisable, novel shapes that are all broadly matched in curvature complexity. Critically, these shapes are all mathematically defined (in terms of sine waves), and through differentiation we can identify points of peak curvature (i.e. ridges and troughs). However, as it is likely that not all ridges and troughs around the contour will be salient, we only counted those that were above a specific size. Specifically, for a given point of peak curvature, we passed a line through the points of inflection on either side of it. As long as the curvature point was greater than  $0.1^\circ$  visual angle above or below that line, it was included.

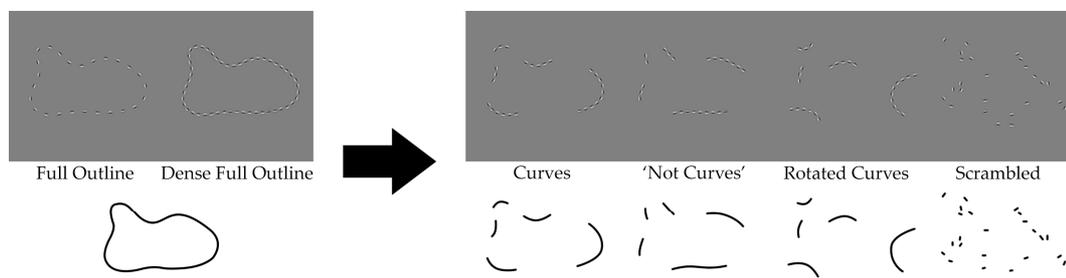
To render the shapes, Gabor elements (small patches of sinusoidally-modulated luminance in a Gaussian aperture) were placed around each shape’s contour. These elements can be perceptually joined into a contiguous contour, however as they are not actually conjoined they can easily be broken apart (destroying coherence whilst preserving low-level cues). This allows us to either render the shape as a whole, as just shape fragments, or we can completely scramble the shape, all whilst maintaining the number of elements depicted. Each Gabor element was  $1/3^\circ$  visual angle in diameter with spatial frequency of 6 cycles per degree, at 50% contrast. The phase of the patch alternated about the contour (only even numbers of Gabor elements were used to ensure that two Gabor elements with the same phase could never be adjacent). This phase alternation has only minor effects on contour integration, but should disrupt integration based upon simple cells (Hess et al., 2003). Gabor elements were filtered with a Gaussian filter (SD  $1/18^\circ$  visual angle). We ensured that the minimum distance between Gabor elements was  $1/3^\circ$  visual angle to avoid overlap, all elements were rendered on a mid-grey background.

The shape stimulus set was rendered with these Gabor elements in six different conditions (Figure 7.7):

1. *Curves*: this was the full shape outline rendered as contour fragments, placed around points of peak curvature. To identify which parts of the contour to include, we took two given points of peak curvature (identified using methods

described above) and quartered the distance between them. The two quarters closest to the two points of peak curvature were included. By repeating this around the whole contour we can select 50% of the total outline to be included as shape fragments.

2. *'Not Curves'*: this was simply the converse of the 'Curves' condition, i.e. the remaining 50% of the contour. Whilst some curvature information was unavoidably present, the contour fragments in this section were generally much straighter. This condition should preserve shape information whilst removing curvature information.
3. *Dense Full Outline*: this was the full outline rendered with Gabor patches. As other conditions are only rendering 50% of the contour, this condition contains twice the number of Gabor elements of any other.
4. *Full Outline*: as 'Dense Full Outline' except that every other Gabor element was deleted (whilst maintaining phase alternations across adjacent elements). This matched the number of Gabor elements in this condition to all other conditions (except 'Dense Full Outline').
5. *Rotated Curves*: as 'Curves' except that every contour fragment was rotated through  $180^\circ$  about that fragment's centroid. This condition should preserve curvature information whilst removing shape information.
6. *Scrambled*: as 'Full Outline' except each Gabor element was spatially jittered  $2/3^\circ$  visual angle at a random angle away from its original position. Note that the minimum  $1/3^\circ$  spacing between Gabor elements was still enforced. This condition should largely preserve low-level cues (e.g. orientation and contrast energy) whilst removing any coherent structure.



**Figure 7.7.** Example Gabor Element Stimuli. Across the top row we see the Gabor element stimuli rendered as they would be for participants (albeit at 100% contrast for visibility), the bottom row shows schematic versions for illustrative purposes.

The average number of Gabor elements in each condition were as follows: 'Curves' 28.24 (SD 3.30), 'Not Curves' 28.00 (SD 3.43), 'Dense Full Outline' 54.00 (SD 5.79), 'Full Outline' 27.00 (SD 2.90), 'Rotated Curves' 28.24 (SD 3.30), 'Scrambled' 27.00 (SD 2.90). This demonstrates that the various conditions should generally be well-matched in terms of low-level visual properties.

### 7.7.2 Data Acquisition and Analysis

The Gabor element stimuli were presented centrally in a block design fMRI experiment (TR = 3000ms, TE = 30ms, voxel size =  $2 \times 2 \times 2 \text{mm}^3$ , flip angle =  $84^\circ$ , matrix size  $96 \times 96 \times 39$ , FOV = 19.2cm), with eight 5min 33sec runs. An additional 9 seconds were prepended to each run to ensure magnetisation reached a steady state (leading to a total run length of 342sec).

A given run was split into three sets of six 15sec blocks (one per stimulus condition), with 21sec baseline (grey screen plus fixation only) after each set. Block ordering was counterbalanced within-participant using a balanced Latin square design repeated four times across the eight runs. This ensures that every condition is equally likely to occur after every other condition. Within a stimulus block, each image was presented on a mid-grey background for 0.4sec, with an inter-stimulus interval of 0.1sec (therefore 30 images per block). Stimulus presentation order was randomised, with every shape occurring twice in each condition per run. In addition, every block contained two sequentially repeated images. Participants maintained fixation whilst performing a one-back task (pressing a button every time a given image was presented twice in succession).

Data were analysed using FEAT (Worsley, 2001). The first three volumes of each run were discarded, the high-pass filter cut-off point was set to 60sec (correcting for low frequency drift), FILM prewhitening and spatial smoothing (Gaussian kernel with FWHM of 4mm; twice voxel size) were used, motion was corrected for (motion parameters were also entered as confound covariates). All six conditions were entered as separate explanatory variables (EVs), convolved with a double-gamma HRF, and contrasts were run to compare each condition to baseline. Runs were combined within-participant using fixed effects analysis with cluster correction ( $Z > 2.3$ ,  $p < .05$ ), all data were retained in the high-resolution structural space.

Percent signal change over baseline was calculated for each ROI in each condition using FEATQuery.

## 7.8 Results – Gabor Element Experiment (Expt. 2)

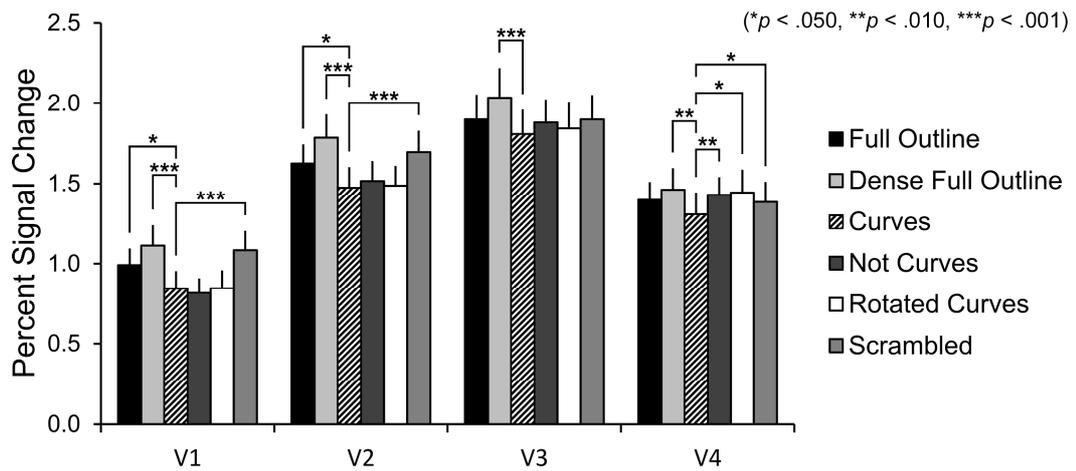
To analyse the results for this experiment we ran repeated-measures ANOVAs comparing percent signal change across the six conditions for each ROI. The Greenhouse-Geisser correction was used when the assumption of sphericity had been violated (as indicated by Mauchly's test) and corrected degrees of freedom are reported.

Planned-comparisons were set-up to compare the 'Curves' condition to all others, as this was the main ('pivotal') condition of interest against which we wanted to evaluate the other conditions. Briefly, in shape-sensitive Lateral Occipital ROIs we expected the two 'Full Outline' conditions to elicit greater activity than 'Curves', due to the more explicit shape information. 'Curves' should elicit greater activation than both 'Not Curves' and 'Rotated Curves', due to the lack of curvature and shape information in those two conditions respectively. The 'Scrambled' condition should elicit minimal activity as it has neither shape nor curvature information.

Starting with earlier visual areas (Figure 7.8), significant main effects of condition were found in V1, V2 and V3 (V1:  $F(5,55) = 15.88, p = 1.2 \times 10^{-9}$ ; V2:  $F(2.79, 30.64) = 25.13, p = .039$ ; V3:  $F(5,55) = 6.59, p = 7.2 \times 10^{-5}$ ). Breaking it down across conditions, V1 and V2 showed identical results profiles with 'Curves' eliciting significantly less activation than the 'Full Outline' (V1:  $p = .018$ ; V2:  $p = .024$ ), 'Dense Full Outline' (V1:  $p = 1.4 \times 10^{-4}$ ; V2:  $p = 1.6 \times 10^{-5}$ ) and 'Scrambled' (V1:  $p = 1.9 \times 10^{-5}$ ; V2:  $p = 2.3 \times 10^{-4}$ ) conditions. V3 showed less differentiation across conditions. 'Dense Full Outline' again elicited significantly greater activity than 'Curves' ( $p = 4.4 \times 10^{-4}$ ), but the 'Full Outline' and 'Scrambled' conditions did not quite reach significantly greater activity ( $p = .053$ ;  $p = .056$  respectively). There was also a hint that 'Curves' might elicit greater activity than 'Not Curves' ( $p = .051$ ).

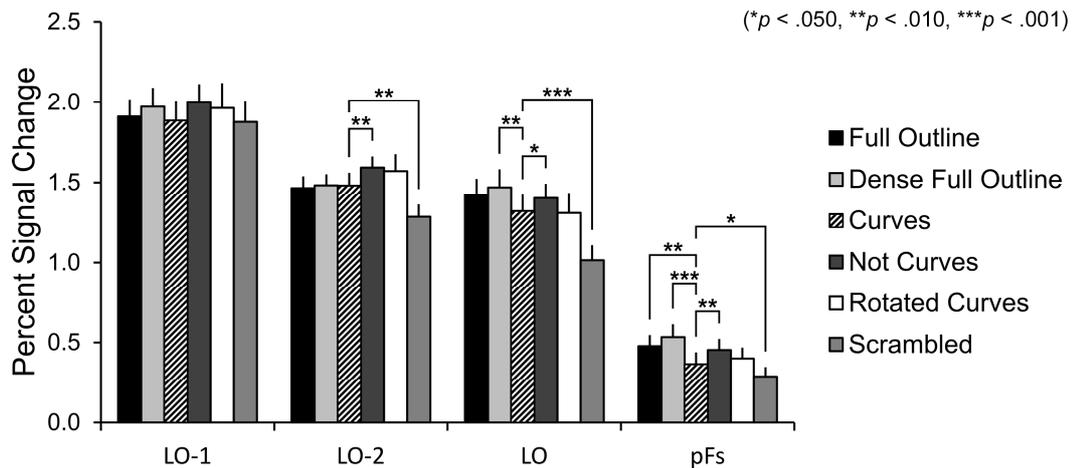
V4 showed a very different pattern of results (Figure 7.8). Again there was a significant main effect of condition ( $F(2.87, 31.57) = 3.04, p = .046$ ), with 'Curves' eliciting significantly less activation than all other conditions apart from 'Full Outline' ('Dense Full Outline':  $p = .004$ ; 'Not Curves':  $p = .010$ ; 'Rotated Curves':  $p = .036$ ; 'Scrambled':  $p = .043$ ).

We also tested the control ROI, VO (combined retinotopically defined VO-1/2 ROIs, as described earlier), which notably showed an identical response pattern to V4 ( $F(2.14,23.49) = 8.81, p = .001$ ; ‘Dense Full Outline’:  $p = .006$ ; ‘Not Curves’:  $p = 1.4 \times 10^{-6}$ ; ‘Rotated Curves’:  $p = .005$ ; ‘Scrambled’:  $p = .011$ ). This could indicate some degree of similarity between these regions, a reasonable proposition given their anatomical proximity.



**Figure 7.8.** Percent signal change across early visual cortex (V1-V3 plus V4) for each condition. Significance reflects the result of planned comparisons comparing the ‘Curves’ condition to all others. Error bars represent the SEM.

Turning to Lateral Occipital ROIs (Figure 7.9), first LO-1 showed no significant main effect of condition ( $F(5,55) = 1.72, p = .146$ ), implying a flat profile of results. Significant main effects were identified in LO-2 ( $F(5,55) = 8.92, p = 3.0 \times 10^{-6}$ ) and LO ( $F(5,55) = 24.28, p = 8.2 \times 10^{-13}$ ). LO-2 is the first region to show significantly less activity for ‘Scrambled’ stimuli versus ‘Curves’ ( $p = .001$ ). Intriguingly, the ‘Not Curves’ condition was the only condition to elicit significantly greater activity than ‘Curves’ ( $p = .007$ ). The same was true for LO, with ‘Not Curves’ eliciting significantly greater activity ( $p = .016$ ) and ‘Scrambled’ eliciting significantly less activity ( $p = 5.1 \times 10^{-5}$ ) than ‘Curves’ respectively. However, LO is also the first Lateral Occipital region to show significantly greater activity for ‘Dense Full Outlines’ than ‘Curves’ ( $p = .009$ ), and a hint of a ‘Full Outline’ preference (although it does not meet significance;  $p = .065$ ). Finally, results within pFs mirror that of LO, with significantly greater activity for ‘Full Outline’ ( $p = .002$ ), ‘Dense Full Outline’ ( $p = 1.3 \times 10^{-4}$ ) and ‘Not Curves’ ( $p = .001$ ) conditions over Curves, and significantly less activity for ‘Scrambled’ ( $p = .031$ ).



**Figure 7.9.** Percent signal change across Lateral Occipital ROIs (LO-1/2, LO) plus pFs for each condition. Significance reflects the result of planned comparisons comparing the ‘Curves’ condition to all others. Error bars represent the SEM.

### 7.8.1 ‘Not Curves’ versus ‘Curves’ – concentricity?

In Lateral Occipital ROIs LO-2 and LO, as well as pFs, we found that the ‘Not Curves’ condition elicited significantly greater activity than the ‘Curves’ condition, contrary to our expectations. One possible reason for this could depend on differences in the ways shape protrusions are rendered in the ‘Curves’ and ‘Not Curves’ conditions. Specifically, for ‘Curves’ the protrusions will likely be rendered at the base, due to the concave curvature where the protrusion meets the ‘base shape’, and at the endpoint due to the convexity where curvature peaks. This should result in a relatively concentric profile. Conversely, in the ‘Not Curves’ condition protrusions will likely be rendered along their lengths, as these will be the straighter segments, resulting in a more radial profile. To test whether this was the case, for every shape in each condition we iterated over all Gabor elements and calculated each element’s angle relative to a line passing through that element, and the shape’s centroid. Smaller angles will reflect more radial Gabor elements, angles tending towards 90° would reflect a more concentric (circular) shape.

Similar average angles were observed for the ‘Dense Full Outline’ (Mean 55.16°, SD 7.69°), ‘Full Outline’ (Mean 55.18°, SD 7.68°) and ‘Scrambled’ (Mean 54.134°, SD 7.27°) conditions. This is to be expected as these conditions all sampled uniformly about the shape’s contour. Slightly more concentric angles were found for ‘Rotated Curves’ (Mean 59.72°, SD 6.58°), and to a greater extent ‘Curves’ (Mean 63.11°, SD 6.46°). However, angles in the ‘Not Curves’ condition were indeed considerably

more radial (Mean  $46.57^\circ$ , SD  $9.93^\circ$ ). A one-way ANOVA demonstrated that the differences between conditions were highly significant ( $F(5,246) = 22.29$ ,  $p = 2.0 \times 10^{-18}$ ). Bonferroni-corrected post-hoc tests revealed that angles in the ‘Not Curves’ condition were significantly lower than those in all other conditions (‘Dense Full Outline’:  $p = 9.3 \times 10^{-6}$ ; ‘Full Outline’:  $p = 8.7 \times 10^{-6}$ ; ‘Curves’:  $p = 2.3 \times 10^{-18}$ ; ‘Rotated Curves’:  $p = 2.1 \times 10^{-12}$ ; ‘Scrambled’:  $p = 8.8 \times 10^{-5}$ ). Additionally, angles in the ‘Curves’ condition were significantly greater than those in the ‘Dense Full Outline’ ( $p = 5.4 \times 10^{-5}$ ), ‘Full Outline’ ( $p = 5.8 \times 10^{-5}$ ) and ‘Scrambled’ ( $p = 5.6 \times 10^{-6}$ ) conditions. ‘Rotated Curves’ angles were significantly greater than angles in the ‘Scrambled’ condition ( $p = .023$ ).

### 7.9 Discussion – Gabor Element Experiment (Expt. 2)

Activation across early visual areas (V1, V2 and to a lesser extent V3) was relatively uniform across the ‘Curves’, ‘Not Curves’ and ‘Rotated Curves’ conditions, which is to be expected given that these conditions are all based around shape fragments with similar path lengths. Strong preferences were found for the ‘Dense Full Outline’ condition, likely as this condition had twice the number of Gabor elements as any other (resulting in greater contrast energy). Slight preferences were also found for the ‘Full Outline’ condition; at least two possibilities could explain this result. First, as the Gabor elements are spread across the whole perimeter (versus isolated contour fragments in other conditions), a larger range of the visual field would be activated during this condition. Naturally, this would stimulate larger portions of the retinotopically-tuned early visual areas leading to greater activation. Alternatively, it is possible that activation in these areas reflects path length. If the Gabor elements could be completed into a cohesive contour, whether from feedback (Shpaner et al., 2013) or not, then the ‘Full Outline’ condition would contain greater path lengths than most other conditions. The former explanation is perhaps more likely, as it would also explain why the ‘Scrambled’ condition elicits greater activity in these regions.

Results in V4 are harder to interpret. Literature on Macaque V4 suggests that it contains some degree of curvature tuning (Pasupathy and Connor, 1999, 2001, 2002; Carlson et al., 2011; Yue et al., 2014), however this would be hard to reconcile with the results presented here. Instead, the ‘Curves’ condition elicits less activity compared to all other conditions but ‘Full Outline’ (notably making this the only

region to show preferential activity for ‘Rotated Curves’ over ‘Curves’). The preferences for ‘Dense Full Outline’ and ‘Scrambled’ conditions could plausibly be due to the same low-level explanations used for earlier visual areas V1 and V2; namely contrast energy and increased spatial coverage. The ‘Not Curves’ preference could reflect greater sensitivity to radial over concentric elements, given that ‘Not Curve’ fragments were generally more radial. Indeed, sensitivity to radial orientations have been observed in V4 (Wilkinson et al., 2000), however the same paper showed sensitivity to concentric stimuli in V4, as did Dumoulin and Hess (2007). An alternative possibility is that V4 is showing preferences for non-shape fragments, which would explain greater activation for both ‘Not Curves’ and ‘Rotated Curves’ over ‘Curves’ (‘Not Curves’ being notably harder to perceive as a cohesive shape than ‘Curves’). One could imagine a mechanism identifying unbound contour fragments, passing them up to higher visual areas (e.g. LO), and then being suppressed by feedback when said contours were successfully ‘bound’ into figure or ground. Such a mechanism could play a useful role in forming a cohesive picture out of the various fragments that form the visual world.

One final consideration is that percent signal change cannot necessarily rule out underlying neural representations. For example, a sparse curvature encoding representation may have small clusters of highly active neurons stimulated during the ‘Curves’ condition, whereas there may be larger regions of less efficient activity for other conditions. In such a scenario, there could actually be greater net activity for conditions that the region is processing sub-optimally. Indeed, greater neural decoding accuracy despite lower overall percent signal change is possible in V1 at least (Kok et al., 2012). As such, follow up studies focusing on representational similarity analysis (i.e. event-related paradigms) are warranted to afford firmer conclusions about neural representations in this and other areas.

Finally, we turn to Lateral Occipital ROIs. The finding of no between-condition differences in LO-1 was surprising. It is possible that LO-1 is simply not tuned to any of the stimuli used, however we note that LO-1 was robustly activated by all conditions (LO-1 percent signal change was close to 2.0% across conditions versus about 1.5% for equivalent Lateral Occipital ROIs). Most research on LO-1 suggests that it plays some role in orientation processing (Larsson et al., 2006; Montaser-Kouhsari et al., 2007; Silson et al., 2013), and so it is likely that LO-1 was

responding to the orientation energy carried by the Gabor elements. If we accept that LO-1 does not discriminate between conditions (as opposed to simply failing to find true significant differences), then our results suggest LO-1 does not preferentially respond either to closed or collinear contours, or sparsely arranged ‘random’ orientations. In the previous experiment detailed in this chapter, we concluded that LO-1 might respond to orientation information regardless of spatial location. Such a mechanism could conceivably account for these findings. All the Gabor elements in each condition originate from the same underlying contours, and so should have roughly equivalent orientation energy. Even though the ‘Dense Full Outline’ condition has twice the number of Gabor elements, those elements are much more densely arranged and so it may not yield much additional orientation information (adjacent elements will have almost identical orientations the majority of the time). As such, abstracted orientation tuning remains a plausible account of LO-1’s response profile, however as with V4 the current experiment cannot make any firm conclusions due to the limitations of a block design.

In LO-2, then LO and pFs, a robust ‘Curves’ over ‘Scrambled’ preference finally emerged, indicating sensitivity to coherent structure and collinear elements. We also noticed a slight (but consistent) trend of greater activity for ‘Not Curves’ over ‘Curves’, contrary to our original hypothesis of shape representations based upon points of peak curvature. This prompted the additional analysis that assessed the angles of Gabor elements across conditions. This analysis identified significantly more radial orientation information in the ‘Not Curves’ condition when compared to any other. In the first experiment detailed in this chapter, we concluded that LO-2 and LO may respond to a ‘Lobe prominence’ dimension, namely the degree to which protrusions actually protrude from the coarser global profile of a shape. The observed ‘Not Curves’ preference in this study could be taken as corroborating evidence for this finding. Shapes with more prominent protrusions would contain more radial energy, as the shape essentially trends away from the perfectly concentric circle. As such this argues for a subtly different representation based not upon points of curvature change, but instead upon radial protrusions from the broader shape outline.

The final point worth noting is that Lateral Occipital ‘Dense Full Outline’ over ‘Curves’ preferences only appear in LO and pFs. Whilst earlier visual areas do show

‘Dense Full Outline’ preferences, this is likely due to the greater contrast energy in this condition. That explanation is unlikely to underlie the observed ‘Dense Full Outline’ preferences in LO and pFs, as the preference is not observed in adjacent Lateral Occipital ROIs LO-1 and LO-2. A preference for the standard ‘Full Outline’ condition does also emerge in pFs, but does not quite reach significance in LO. To our knowledge this is the first observed dissociation between LO-2 and LO, but interpretation is difficult. The most plausible account is that shape sensitivity only truly emerges in LO, with LO-2 instead being a ‘precursor’ or preliminary shape processing area. This would explain why activation elicited by standard ‘object over scrambled object’ LOC localisers typically only overlaps with LO-2, rather than stimulating it completely (Sayres and Grill-Spector, 2008; Silson et al., 2016; Vernon et al., 2016).

To conclude this section, results in early visual areas (V1-V3) were uncontroversial. We found greater activity in conditions with more Gabor elements (‘Dense Full Outline’) and conditions in which the Gabor elements were presented more sparsely (‘Full Outline’, ‘Scrambled’). Results in V4 and LO-1 were harder to interpret. V4 could plausibly be showing similar patterns of activity to that observed in earlier visual areas, yet the ‘Not Curves’ and ‘Rotated Curves’ preferences hint at additional influences here (perhaps sensitivity to unbound contour fragments). Activity within LO-1 was robust, however it showed no hints of shape sensitivity (whether explicit shapes for full outlines, or implied shapes for ‘Curves’ or ‘Not Curves’) and instead may just be showing general activation for the orientation energy contained in Gabor elements. Preferences for coherent form emerges in LO-2 onwards, with ‘Scrambled’ shapes eliciting less activity than ‘Curves’. Intriguingly, a ‘Not Curves’ preference was also observed across LO-2, LO and pFs. This may reflect sensitivity to radial elements; specifically, protrusions that extend away from the shape’s centroid. Finally, we noted a further sensitivity for full shape outlines emerging around LO and pFs. This could indicate that true shape processing only really emerges in these later regions, with LO-2 instead being more of a preliminary processing area.

## 7.10 General Discussion

The aim of these two experiments was to explore the ‘Shape Complexity’ dimension that was found to explain significant amounts of variance in shape-sensitive Lateral Occipital Cortex. ‘Shape Complexity’ appeared to capture the amount of detail

around a shape's contour, and so we speculated that it was some representation based upon curvature content. Specifically, we hypothesised that shape selective areas might process those shapes based around points of maximal curvature change. This could either be the number of points of curvature change, or magnitude of those points of curvature change.

Surprisingly, our current experiments argue against that hypothesis. The first experiment, using radial frequency (RF) patterns, suggests that representations in LO are not based upon the number of points of curvature change. 'Frequency', a variable representing the number of lobes (or 'protrusions'), only correlated with neural similarity in LO-1 suggesting later regions (LO-2, LO) were not sensitive to this manipulation. It is possible that the stimulus set's artificiality played some role here. All protrusions in a given pattern were identical, perhaps placing emphasis on the nature of the protrusion rather than the number of them. Nevertheless, we have no compelling evidence to suggest that either LO-2 or LO responds to frequency, at least over the (admittedly limited) frequency range tested here.

Instead, we found relatively robust correlations between amplitude (size of a lobe/protrusion) and neural similarity across Lateral Occipital cortex. This would imply that it *is* the magnitude of curvature change points that matter. However, the finding was obfuscated to some extent by the later evidence for two discrete influences across Lateral Occipital extrastriate regions. Namely, we found a 'Lobe prominence' dimension that correlated primarily with LO-2 and LO; this factor likely represents amplitude as it was originally intended (i.e. representing the degree to which protrusions actually protrude). Secondly, we found a 'Lobe curvature' representation that likely does represent some aspect of curvature information; whether lobes were sharp/acute, or more rounded. This again correlated well with LO-2 and LO, but now also LO-1 and to a less extent V4. Perhaps a good way to think of these separate influences is to imagine a Gaussian curve. To define such a curve, you need both a height (reflecting 'Lobe prominence') and a 'width' or standard deviation (reflecting 'Lobe curvature'), and these two values can be manipulated independently.

On first impressions, our second experiment did not serve to refine our understanding of these dimensions. Instead we found a profile of results that

contrasted considerably with our original expectations. Primarily, we found that shapes represented purely by points of maximal curvature change ('Curves') evoked less neural activity than the converse condition; shapes represented by straighter segments ('Not Curves'). Whilst this cannot necessarily rule out representations based on curvature change for design choices discussed previously (namely the use of a block design), it does undermine that hypothesis to some extent. However, additional analysis prompted by these results showed that shapes in the 'Not Curve' condition generally contained significantly more radial information than shapes in the 'Curves' condition. That is, 'Not Curve' shapes contained more Gabor elements oriented radially with respect to the shape's centroid. Based upon this, and the previous study's results, we argued for a representation based primarily upon shape protrusions.

Specifically, we are suggesting that in a given shape there will 'base shape' (akin to the low complexity shapes in Vernon et al., 2016) that could be considered a coarse (global) low-pass filtered version of the overall shape outline. Situated on this base shape are shape features that emerge radially (protrusions), and we suggest that Lateral Occipital regions may be tuned to these features. The tuning appears to be for at least two (independent) characteristics; the degree to which a protrusion actually protrudes and the width or acuteness of that protrusion. The exact roles (if any) that LO-1 and LO-2 play in terms of LO's representation are currently unclear. However, we now have evidence for dissociations both between LO-1 and LO-2 (mirroring earlier work; Silson et al., 2013) and now LO-2 and LO. If these regions are all involved in an object-processing stream, then this suggests that these regions are in part discrete processing units.

## Chapter 8. General Discussion

### 8.1 Overview

The central aim of this thesis has been to investigate the roles of the retinotopically-defined regions of interest (ROIs) LO-1 and LO-2 (Larsson and Heeger, 2006), and how those roles relate to the functional properties of object-selective Lateral Occipital Complex (LOC) (Malach et al., 1995; Grill-Spector et al., 2001). Whilst there is an extensive body of literature detailing the properties of the LOC, in those studies the LOC is often explored in isolation (sometimes with V1 as a ‘low-level control’) (e.g. Kourtzi et al., 2003; Stanley and Rubin, 2003; Eger et al., 2008; Haushofer et al., 2008a; Andresen et al., 2009; Snow et al., 2011). However, the object tuning in LOC cannot arise in isolation; it must be derived from its inputs in earlier retinotopic visual areas. As such, by exploring responses to shape across visual cortex, we aimed to characterise the tuning within LOC from a bottom-up perspective. LO-1 and LO-2 were of particular interest in this regard, as these regions show some degree of overlap with activation elicited by the classic ‘objects > scrambled’ LOC localiser (Larsson and Heeger, 2006; Sayres and Grill-Spector, 2008; Silson et al., 2016). Furthermore, the fact that they are complete hemifield maps and are functionally dissociable to some degree (Silson et al., 2013) implies that they may be discrete ‘processing units’. As such, we typically discuss these regions as if they may be preliminary shape processing stages for the LOC. We explored the validity of this claim, plus more generally the functional properties of these ‘mid-level’ visual areas in a series of six fMRI experiments. The first two studies conducted were specifically aimed at testing the response characteristics of LO-1 and LO-2, whereas the latter four experiments focused on the underlying representations within those visual field maps, and how those representations related to those in LOC.

### 8.2 Summary of Work

First, we will briefly summarise the main findings from each study. Implications for our key visual field maps (LO-1, LO-2 and LO) will follow.

Our first experiment aimed to further explore the double dissociation identified by Silson et al. (2013). Specifically, Silson et al. found that transcranial magnetic stimulation (TMS) applied to LO-1 (but not LO-2) disrupted orientation judgments

of sinusoidal gratings, whereas TMS applied to LO-2 (but not LO-1) disrupted shape curvature judgments of a radial frequency (RF) pattern (Wilkinson et al., 1998).

Whilst compelling, there are a large number of differences between sinusoidal gratings and RF patterns. This means that it is difficult to determine exactly what aspect of Silson et al.'s stimuli each ROI is causally involved in. Due to this, we ran an fMRI study that compared neural responses to RF pattern-based orientation and shape judgments. Specifically, across separate blocks participants determined either which of two bilaterally presented RF patterns was most clockwise (an orientation judgement task), or which was spikier (a shape curvature processing task). LO-2 generally showed greater activation when compared to LO-1, with no differences in activation across tasks, implying it was generally activated by shape stimuli.

Intriguingly, LO-1 was the only region in which we identified a dissociation across tasks, and it was in the opposite direction to our hypotheses. We found that LO-1 showed significantly greater activation for the shape task over the orientation task. This was also reflected in behavioural data; those with generally greater LO-1 activation tended to have better performance on the shape task.

Our second study tested sensitivity to second-order shape stimuli (i.e. shapes defined from cues other than luminance). We noted that earlier work tended to find second-order sensitivity in LO-1, but not (or at least to a limited extent) in LO-2 (Larsson et al., 2006; Montaser-Kouhsari et al., 2007), despite known second-order sensitivity in LO (Grill-Spector et al., 1998a; Murray et al., 2002a; Ritzl et al., 2003; Stanley and Rubin, 2003). This raised questions about the relationship between LO-2 and LO. However, the studies finding second-order sensitivity in LO-1 used edge or grating stimuli, which we know LO-2 is less sensitive to (Silson et al., 2013). As such, we hypothesised that it would be LO-2, and not LO-1, that would show greatest second-order sensitivity when those second-order cues defined shapes. This prediction largely held true. LO-2 was the first region to show robust activation for our second-order shape stimuli, mirroring activation in later LO (and to a lesser extent pFs). LO-1 did show significant preferences for second-order shapes at random versus fixed orientations (implying a release from adaptation and hence second-order sensitivity), but generally its responses were low and did not significantly differ from baseline.

The next set of two studies (Vernon et al., 2016) shifted focus, moving away from pooled percent signal change across ROIs towards more nuanced explorations of how our stimuli were *represented* within an ROI. We used two discrete stimulus sets; ‘standard’ (or ‘recognisable’) and ‘scrambled’, which were organised along dimensions of exemplar (bird, cat, rabbit) and Fourier descriptor (FD) (Zahn and Roskies, 1972) content (low, medium and high levels of detail about contour). A rapid event-related fMRI study probed the neural responses unique to each stimulus, allowing us to generate neural similarity matrices for each ROI. These detail, for a given ROI, the shapes that elicit more similar patterns of distributed activity. The similarity matrices could then be compared to a variety of ‘predictors’, allowing us to interpret the nature of the shape representations across visual cortex. Critically, two discrete organisational principles emerged. The first (‘Shape Profile’) captured substantial variance in earlier visual areas V1-V3 and to a lesser extent V4, so likely reflected retinotopic responses. Of more interest was the second dimension, ‘Shape Complexity’. This captured considerable variance not only in LO and (to a lesser extent) pFs, but also in LO-2. Essentially, this suggests that LO-2 and LO share very similar shape representations, despite the fact that the regions are retinotopically and functionally defined respectively. Furthermore, LO-1 showed a mixed representation, with both Shape Profile and Shape Complexity dimensions explaining some variance in this region.

We followed up this study with a brief behavioural experiment, asking participants to perform midpoint judgements (psychophysical method of bisection) on shapes that had been manipulated in summed Fourier amplitude content. We found that participants’ judgements corresponded almost linearly to summed amplitude content, suggesting that the manipulations of FD content in previous work (Vernon et al., 2016) were capturing perceptually relevant details. Furthermore, participants’ judgements were comparable across both standard and Fourier phase-scrambled stimuli, again implying that this was a valid manipulation to control for semantic content whilst preserving contour complexity.

Finally, our last two studies were intended to further explore the ‘Shape Complexity’ dimension which captured variance in LO-2 and LO. The first study used a set of RF patterns differing in orthogonal dimensions of amplitude and frequency. We again used a rapid-event related paradigm, with the intent of exploring neural

representations across visual cortex. We found that after controlling for stimulus size, the only area to correlate with frequency was LO-1, whereas amplitude correlated with LO-1, LO-2, LO and to a lesser extent pFs. However, the amplitude result was slightly more involved, as we could again break it down into two discrete dimensions when combined with other predictors. The first dimension, ‘lobe prominence’, appeared to capture the degree to which lobes protruded, and explained variance most consistently in LO-2 and LO. The second dimension, ‘lobe curvature’, appeared to be related to the curvature of a lobe (e.g. narrow, acute curvature versus broader, more shallow curvature). This dimension captured variance across LO-1, LO-2 and LO.

The second in this set of studies assessed curvature tuning more specifically. We used a set of shape represented either by a full outline, points of curvature change only (‘curves’), the converse of those points (‘not curves’), the curvature change points rotated through 180° (‘rotated curves’) or a completely scrambled outline. In this study, we expected ‘curves’ to elicit greater activation than all but the full outline conditions, predicated upon the theory that LO’s shape representation may in part be based upon curvature information. However, this hypothesis was largely contradicted. We actually found a ‘not curves’ preference emerging in LO-2, and this continued through to LO and pFs. Full outline preferences only emerged in LO and pFs. All three regions did at least respond more to curves than scrambled stimuli, indicating sensitivity to coherent form. LO-1 showed no distinctions between our various conditions, responding relatively strongly to all.

### **8.3 Interpretation of the Results**

We will now collate and consider all results described in this thesis across our key regions of interest; LO-1, LO-2 and LO.

#### **8.3.1 LO-1**

The first region to examine, LO-1, is perhaps the most difficult to interpret. At the start of this PhD we largely saw LO-1 as a discrete visual region that played some role in orientation processing (Larsson et al., 2006; Montaser-Kouhsari et al., 2007; Silson et al., 2013). However, the results presented here obfuscate this to some extent. Most surprising is the apparent contradiction between the double dissociation identified by Silson et al., and our finding of LO-1’s shape sensitivity in the first

empirical chapter (Chapter 3). Namely, Silson et al. found that LO-1 was causally involved in an orientation judgement task, not a shape-based task, yet our results showed that LO-1 responded more to the shape specific task compared to the (shape-based) orientation task. There are a number of differences between our orientation task, and that used by Silson et al., so it is difficult to determine exactly why LO-1 responded more weakly in this task compared to its ‘non-preferred’ task. It could indicate that there is something specific to lines or edges that LO-1 is attuned to. However, our shape-processing task was in principle very similar to that used by Silson et al.; both tasks used RF amplitude detection yet we found LO-1 activation for our task whereas Silson et al. found no evidence for LO-1 involvement. The key differences between our task and Silson et al.’s were that we used bilateral presentation, with relatively small orientation differences between reference and test stimuli (versus serial presentation and larger orientation differences respectively in Silson et al.).

We speculated that LO-1 might have processed *local* orientation cues that could have been used to solve the shape-related task. Changes in an RF pattern’s amplitude will cause predictable changes in the range of orientations around the contour (namely, amplitude increases will decrease concentricity). Bilateral presentation may have encouraged participants to attend to these orientation cues as a means to quickly evaluate stimuli. With serial presentation, it would likely be easier to compare the global form of the shape across exemplars (perhaps because the participant does not have to split attention across two stimuli simultaneously).

The theory that LO-1 processes local orientation cues is largely consistent with our remaining studies. First, we note that LO-1 did respond significantly more to second-order RF pattern stimuli at random orientations, when compared to fixed orientations. This is intriguing, given that we have just discussed LO-1’s limited responses during an RF pattern-based orientation task. However, the study examining responses to second-order stimuli did not *necessitate* bilateral attention (despite bilateral presentation). As such, the changing orientations of the RF patterns could have been processed on a local level. Naturally, a set of RF patterns at random orientations will contain considerably more orientation energy than a static RF pattern across a block, and they so would be expected to elicit greater activation in a region tuned for orientation information.

Of greater relevance is our study that directly compared neural tunings to properties of RF stimuli across visual cortex (Chapter 7). First, the ‘lobe curvature’ dimension we identified in this study correlated well with neural similarity in LO-1. We suggested that this dimension captured aspects of an RF pattern’s lobes; how smooth/rounded or sharp/acute they are. This fits with the theory that LO-1 responded significantly to the RF pattern amplitude task due to changes in local orientation cues, as such cues will also be informative about the curvature of an RF pattern’s lobe. More importantly, we found LO-1 was the only region which would respond to the frequency of an RF pattern. This is particularly noteworthy given that again our RF stimuli were presented at random orientations; blurring any systematic links between orientation and spatial location. This likely explains why earlier visual areas such as V1 (with known orientation tunings; Hubel and Wiesel, 1959) did not show frequency tuning, despite the greater orientation (and contrast) energy included as frequency increases. From this, we can infer that LO-1’s frequency tuning may in part be abstracted from spatial location.

In a sense, this also fits with the idea that LO-2 contains a more ‘abstracted’ shape representation; namely that of Shape-Complexity (Chapter 5). Given how late LO-1 and LO-2 are in the visual processing stream, it is possible that the role of both these regions is to start abstracting the core visual properties in a scene from specific spatial location. Such a feature would be undeniably useful as spatial location is more a feature to be controlled for, than a static property of the visual world. For example, we can still recognise a given object regardless of whether it was presented in the upper left or lower right visual field; some level of abstraction is necessary. This raises the question of what exactly LO-1 is abstracting (if anything)?

We have evidence that LO-1 is less sensitive to shape information when compared to LO-2. For example, in our Gabor element experiment (Chapter 7) LO-1 failed to distinguish between any of the conditions. That is, LO-1’s responses were as strong to completely scrambled Gabor elements as to contour fragments or even a full shape outline. This suggests that LO-1 is less sensitive to coherent structure, perhaps responding to the individual elements rather than contours as a cohesive whole. As such, in the aforementioned studies LO-1 was likely responding to features within a shape rather than the shape *per se* (likely also indicating why LO-1 was found to not play a *causal* role in shape processing tasks in Silson et al.). However, we also note

that in our study comparing shapes across different complexity levels (Chapter 5), LO-1 showed a mixed representation with hints of the ‘Shape Complexity’ dimension emerging. This implies that LO-1 is sensitive to the complexity of an image, although not necessarily the complexity of a shape as we predicted for LO-2 and LO.

One possibility is that LO-1 acts as a more general orientation filter to identify potential regions of interest in a visual scene. To take a scene for example, you could broadly split it into sky and ground based upon frequency information; the sky would generally contain lower frequency details and so could be classed as less salient. The ground could then be further segmented into regions of high orientation information, or entropy (e.g. bushes), and lower orientation information (e.g. roads). Furthermore, regions of coherent orientation information (e.g. parallel, radial or concentric orientations) might be flagged as they could identify regions of structure (i.e. objects; see later discussion). As such, a crude ‘orientation energy’ filter could prove useful as a first-pass segmentation of a visual scene. Whilst it could be argued that earlier visual areas, even V1, could play this role, the proposed abstraction within LO-1 may come into play simply because of the dynamic nature of the world. We have discussed evidence that implies LO-1 is sensitive to the frequency of an RF pattern regardless of its orientation. It would be interesting to test whether such sensitivity remains after alternative affine transformations (scaling, translation etc.). If so, LO-1 may be tracking orientation across the visual world.

### **8.3.2 LO-2**

Our results for LO-2 are more straightforward to interpret, but perhaps no less surprising. The first two studies largely just cemented LO-2 as the more shape sensitive region out of the LO-1/LO-2 pairing. LO-2 generally responded more to our RF patterns when compared to LO-1, regardless of whether participants were performing an orientation- or shape-based judgment task. This was slightly unexpected, given that we knew LO-2 played a causal role in a similar shape processing task (Silson et al., 2013), however this of course does not imply that a (shape-based) orientation task would be less salient for LO-2. We should note though, that for this study both orientation and shape (specifically, shape amplitude) changed simultaneously; the only difference between blocks being the feature that participants attended to. This implies that LO-2 could potentially have been

responding to the amplitude change regardless of which feature was being attended to. However, we doubt this is the case, as activation was greater (but not significantly so) during the orientation task. The opposite would have been expected if LO-2 was only responding to amplitude information.

Our second study identified LO-2 as the first region in the visual hierarchy in which robust sensitivity for second-order shape information emerges. This study was largely confirmatory; there is good evidence for second-order sensitivity in LO (Grill-Spector et al., 1998a; Murray et al., 2002a; Ritzl et al., 2003; Stanley and Rubin, 2003), and so if LO-2 lacked such sensitivity it would raise questions about the LO-2/LO relationship. However, it is unlikely that general second-order sensitivity arises here. This work was based in part upon previous evidence showing good second-order sensitivity to lines or edges in LO-1, but not (or to a lesser extent) in LO-2 (Larsson et al., 2006; Montaser-Kouhsari et al., 2007). As such, it is likely that second-order boundaries are detected earlier in the visual system, and then if those boundaries happen to form shape contours LO-2 will respond. This again implies that LO-2 may be filtering input from earlier regions in the visual system, before passing shape specific input through to LO.

The most compelling evidence for an intrinsic relationship between LO-2 and LO comes from our work suggesting that both regions share the same ‘Shape-complexity’ representation. This result was echoed in our follow-up work exploring the neural representations of RF patterns; both the ‘lobe prominence’ and ‘lobe curvature’ dimensions correlated with LO-2 and LO to similar extents. The nature of LO-2 and LO’s neural representation will be discussed in greater detail below, however the general finding that these regions show such similar neural representations is surprising for two reasons. First, LO-2 and LO are defined in fundamentally different ways; LO-2 with retinotopic mapping, LO based upon functional selectivity. Second, the fact that LO-2 is retinotopically-defined implies that its response tunings should largely reflect the spatial extent of our stimuli. However, the ‘Shape complexity’ dimension in particular showed no hints of a retinotopic basis, yet captured considerable variance in LO-2. The main conclusion that we drew from these results was that LO-2 might be exploiting its large receptive fields (Dumoulin and Wandell, 2008; Amano et al., 2009) to sample across the visual scene; selectively responding to whatever shape features it is attuned to. As

discussed previously for LO-1, implicit in this idea is the theory that the role of LO-1 and LO-2 may be to ‘abstract’ retinotopic input and respond to visual features regardless of spatial location. However, such a theory needs explicit testing. In our studies we typically present stimuli foveally, at relatively constant sizes and a ‘flat’ perspective (i.e. no skew). This means that we are limited in the conclusions that we can draw about LO-2’s abstraction. Nevertheless, we found that LO-2 would respond more similarly to stimuli with such spatially-distinct outlines as e.g. a high complexity bird and high complexity cat, versus e.g. the spatially overlapping high and mid complexity birds. Given this, it would not be surprising to find that LO-2 possesses some invariance across simple affine transformations.

The above makes a strong case for LO-2 as a preliminary processing region for the LOC, however to be classed as *preliminary* there should be some functional differences between the two regions. That is, there should be a level of processing sophistication in LO that is not present in earlier LO-2. Identifying such a dissociation was one of the secondary goals of this thesis, and a few hints emerged in our latter two studies. First, there is a dissociation that was not really discussed in our study comparing neural representations for RF patterns (Chapter 7). One of the earlier steps we took in that study was to control for the size of our stimuli, by regressing out stimulus perimeter and area. However, we first checked for any influence of these parameters across visual cortex. As would be predicted, they were mostly represented in earlier visual areas V1-V3, however both perimeter and area did correlate to some degree with LO-2 (although the relationship was only significant for area). Critically, there was no hint of an area or perimeter correlation in LO. Given that in all other representational similarity analyses LO-2 and LO have shown essentially equivalent response profiles, this is a notable finding. Naturally, the most parsimonious explanation is simply that this is a reflection of LO-2’s retinotopic response profile, but the fact that this is not carried through to LO again implies that LO-2 may be acting as an intermediary between retinotopic and more abstracted responses to shape.

The second, perhaps more intriguing hint of a dissociation comes from our study exploring responses to shapes defined e.g. either as a full outline, points of curvature or ‘not curves’ etc. In this work, we noted that ‘full outline’ preferences only emerged in LO and pFs, with LO-2 showing flat results across the ‘dense full

outline’, ‘full outline’ and ‘curves’ stimulus conditions. This is a difficult result to interpret without additional evidence, however it may imply that LO-2 is not responding specifically to shapes as a whole but rather segments within those shapes. In all other aspects, LO-2 and LO seem to share the same neural representations, as such it would follow that the segments LO-2 responds to are those that are critically informative for later representations. Given the emerging ‘not curves’ preference in LO-2, we suspect that part of LO-2’s role may be to extract radial orientations (as the ‘not curve’ condition contained more radial orientations than all others). Whilst the nature of the shape representation in LO will be discussed in more detail below, briefly we suspect a ‘shape protrusion’ representation may exist. In such a scenario, LO-2 may identify ‘spokes’ about a shapes boundary that cue identifiable features, for example legs or ears etc. on an animal. These features could then be bound into a cohesive shape later in the visual stream.

### 8.3.3 LO

Throughout this thesis, LO (the posterior, dorsal component of LOC) has largely been treated as a ‘foil’ against which we have compared LO-1 and LO-2. Nevertheless, our work has provided considerable insights into the nature of the shape representation within this region.

Early work evaluated neural representations within LO with respect to perceptual shape similarity. That is, the degree to which an observer would behaviourally rate two shapes as having a similar appearance. Two such papers were published in 2008, with slightly contradictory conclusions. Op de Beeck, Torfs & Wagemans (2008b) found representations within both LO and pFs corresponded to perceptual shape similarity, whereas Haushofer, Livingstone & Kanwisher (2008b) found perceptual shape similarity only accounted for neural similarity within pFs (LO being better captured by a metric of physical shape similarity). However, a possible reinterpretation of these results is that *feature similarity* was driving neural similarity within LO across both studies. Op de Beeck et al. found that their perceptual similarity measure was related to shared shape properties. For example, two shapes with ‘spikey’ protrusions would both be rated as more perceptually similar, and generate more similar patterns of neural response in object-selective cortex, than two shapes that differed in the nature of their protrusions. For Haushofer et al., the physical metric that captured responses within LO was related to the similarity

inherent in a linear transition across shapes defined in an aspect ratio and skew shape space. Therefore, by definition two shapes that were closer in this space (and so described as more physically similar) would have more similar features. We also note that a later study explicitly concluded that LO contained a coding for shape features, based upon amplitude manipulations of composite RF-style stimuli (Drucker and Aguirre, 2009).

In our own work, we tested the salience of shape features by manipulating the contour complexity of simple shape outlines. This led to the identification of a ‘Shape complexity’ dimension that corresponded well with neural similarity across LO (and LO-2). This Shape complexity dimension was a composite of multiple predictors that attempted to capture the level of detail in a shape’s contour. Namely the number of ridges and troughs around the contour, number of Fourier descriptors needed to create the shape, compactness (deviation from circularity) and deviation from the shape’s convex hull. In sum therefore, we found shapes with similar levels of detail in the contour will elicit similar patterns of neural response across object-selective cortex. First, this implies a non-retinotopic response profile, as shapes with similar complexity differed considerably in spatial extent. It also reinforces the importance of shape features, as most increases in shape complexity will correspond to the introduction of a shape feature (e.g. the ears emerge on a rabbit).

To briefly elaborate on the notion of a non-retinotopic response profile, we are not suggesting that LO discards retinotopic information entirely. Our work predominantly uses foveally-presented stimuli encompassing a relatively small proportion of the visual field (although cortical magnification means our stimuli may encompass larger parts of visual cortex). As such, we likely bias tunings away from a retinotopic representation towards the more abstract representations identified in previous chapters. If our stimuli varied in spatial position, then such positional information would likely capture variance within LO as others have found (Grill-Spector et al., 1999; Schwarzlose et al., 2008; Kravitz et al., 2010). Indeed, retinotopic biases are well established in this region; most notably preferences for central over peripheral stimuli, plus a general contralateral lower visual field bias (Niemeier et al., 2005; Sayres and Grill-Spector, 2008; Silson et al., 2016). This could indicate hitherto unidentified retinotopic maps (perhaps due to limitations in spatial resolution or stimulus choice), or instead it may be a comorbid response to

the nature of the processing in this region. Objects will typically appear in the lower visual field, and so connections between lower visual field retinotopic input and Lateral Occipital neurons will likely be strengthened. This could potentially manifest itself as a lower visual field bias when Lateral Occipital Cortex is probed retinotopically. Either way, our work suggests that shape information is abstracted from retinotopic input to a certain extent, and we would still argue that abstraction from retinotopy is ultimately necessary to allow position-invariant object recognition (wherever in the visual hierarchy this may occur).

Turning back to our ‘shape complexity’ dimension, we further explored the nature of this representation in two follow up studies. The first used radial frequency (RF) patterns that varied in the size and number of lobes around the pattern’s contour. This essentially asked whether it was the number of shape features, or size of those features, which were most instrumental for LO’s neural representation. The results implied that it was the size of features that was most influential, rather than number. However, RF lobe properties could be further broken down into two discrete factors. The first appeared to capture ‘lobe prominence’; the degree to which an RF pattern’s lobe protruded from the more global circular shape. The second, ‘lobe curvature’, instead captured the ‘breadth’ of the lobe; how acute or broad the curvature was. For an analogy, consider a Gaussian curve; this can be described both by its height (akin to lobe prominence) and a measure of spread (e.g. FWHM; akin to lobe curvature). Both factors were significant predictors of neural similarity within LO (and again LO-2). One caveat with these results, is that they highlight LO’s sensitivity to the specific size and shape of a lobe. However, the nature of an RF pattern means all lobes have the exact same shape about the contour. This may have biased our results away from a frequency representation, and it is possible that LO would be more sensitive to lobe frequency (i.e. number of features), if those lobes/features differed to some extent.

Our second follow-up specifically tested the hypothesis that LO’s underlying representation was based upon points of maximal curvature (regardless of whether that curvature was concave or convex). Specifically, we rendered shapes using contour fragments, highlighting either those maximal curvature points, the converse of those points (‘not curves’), or various alternatives acting as controls. However, we found that it was the ‘not curves’ condition that elicited greater activation in LO, not

the points of maximal curvature. In truth, this study was intended to be largely confirmatory and so the design was not optimal for exploration when the hypothesis was not met. Our previous results (Vernon et al., 2016), the large body of behavioural literature suggesting the importance of points of maximal curvature (Bertamini and Wagemans, 2013) and Macaque literature implying curvature tunings in Inferior Temporal Cortex (Pasupathy and Connor, 1999, 2001, 2002; Connor et al., 2007; Yue et al., 2014) all seemed to imply a representation based around points of maximal curvature. There is some behavioural evidence suggesting straighter segments may improve identification (Panis et al., 2008), however that tended to hold for more complex shapes, whereas our shapes were specifically constrained to simpler outlines. Therefore, this was generally an unexpected result.

It would still be premature to rule out curvature-based representations from this work alone. This study used a block design and as previously discussed, a block design is not optimal for probing underlying neural representations, so we are hesitant to draw firm conclusions from this result. For example, it may be that points of maximal curvature can be represented with small sets of efficient tunings, whereas ‘not curve’ shapes need to recruit a wider population of less efficient tunings. In such a scenario, there may be greater net activity for ‘not curves’, despite them not being the optimal stimulus type for the region. One may also need to consider adaptation. If full outline shapes are indeed represented based upon points of maximal curvature, then the same sets of neurons will be activated for the two full outline conditions and the curves condition, whereas a separate population may be recruited for ‘not curves’. This could potentially have caused the small increase in percent signal change observed for the ‘not curves’ condition. However, the large number of discrete shapes used for this study make this interpretation unlikely.

Despite the above considerations, we cannot discount the fact that the same ‘not curves’ over ‘curves’ preference was observed across three discrete regions (LO-2, LO & pFs). We also note that it was consistent across participants; at least 9 of 12 participants showed the ‘not curves’ > curves preference across the LO-2, LO and pFs ROIs. Therefore, we should not dismiss the result entirely. In a post-hoc follow up, we noted that one key difference between conditions is the amount of ‘radial information’ contained in the Gabor elements; the ‘not curves’ condition having significantly more radial orientations than all others. That is, the Gabor elements in

the ‘not curves’ condition tended to align with imaginary spokes protruding from the shape’s centroid, as opposed to lying orthogonal to them as would be the case with more concentric orientations. Consequently, we suggested that shape representations in Lateral Occipital cortex may be based in part upon shape protrusions; radial projections that emerge from a more global ‘base shape’ outline.

Note that by ‘protrusion’ we do not necessarily imply a point of peak curvature. For example, a chimney on a house could be considered a protrusion, but there is no curvature at its apex. Nevertheless, to extend the previous analogy that chimney could still be fitted with a Gaussian and so have an associated height (‘prominence’) and width (‘curvature’) in line with the two dimensions from our radial frequency work. As such, a representation based upon protrusions is subtly different from one based upon points of peak curvature.

A representation based around protrusions does make some conceptual sense. We can consider shapes from an almost evolutionary, ‘survival of the fittest’ perspective. A perfectly concentric shape is the simplest possible shape - a circle. Deviations from this concentricity (i.e. radial protrusions) will increase complexity and introduce weak points in the object where that protrusion meets the base shape. This means that there should be a cost associated with radial protrusions and therefore, if they exist, they are likely to produce some benefit to that object. Essentially, the protrusions in a shape should be the most salient and important points on that shape’s contour, meaning a representation based on those points is likely based on the most informationally dense part of the object. Lots of shapes will have the same broad global profile, but few will contain the same protrusions in the same locations, and in cases when protrusions are shared, it may indicate important ties between those sets of shapes (e.g. animals share ‘leg protrusions’).

The focus on protrusions also implies a conceptual shift in the way we have currently discussed shapes. In Vernon et al. (2016), the shapes were described with Fourier descriptors (FDs) and manipulated accordingly. FDs capture information about the contour of a shape, and essentially treat the perimeter as a ‘whole’. For example, altering the phase or magnitude of a single FD will elicit changes across the entire perimeter (as demonstrated by the drastic results yielded when rotating the phase of just one FD, to create the scrambled shapes in Chapter 5). Whilst useful, this does

not necessarily reflect the way we actually perceive shapes; although we may see (e.g.) a cat as a complete contour, we can also break it down into constituent parts (ears, tail etc.). Furthermore, small changes such as moving a tail can have profound effects on the external contour (and therefore FDs), without perceptually altering the ‘core shape’.

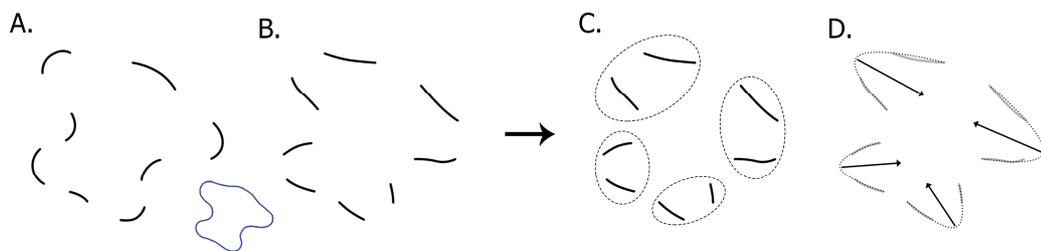
An alternative way to think about shapes (and potentially manipulate them) is to examine their internal structure, rather than contours alone. One way of achieving this is through shape skeletons; the set of internal points that have at least two closest points on the shape’s perimeter (Blum, 1973). This can be calculated with the grassfire transform; by conceptually ‘setting fire’ to the outer perimeter of the shape and marking all points where the fire wavefronts meet, the resultant set of lines forms the skeleton. There is some behavioural evidence that people are unconsciously sensitive to such skeleton information (Firestone and Scholl, 2014). This suggests that shape skeletons may be useful in capturing perceptually salient shape parts, even if only in a conceptual sense. For example, the longest path in the shape skeleton could be thought of as the baseline shape, with branches corresponding broadly to protrusions. With such a framework, a moving tail can affect just one branch point, rather than the shape as a whole. In sum, experiments based around shape skeletons may prove a useful means to further interrogate neural representations across Lateral Occipital Cortex.

#### **8.4 Considering a protrusion-based representation**

If neural representations are based upon shape protrusions, this raises questions about the nature of the ‘shape map’ in LO, if such a map does exist. So far our approach has been largely to test neural similarity in LO against a scattershot of various stimulus properties to see which may prove influential. However, ultimately something more systematic (akin to polar angle or eccentricity maps) will be needed to truly understand human shape representations. If the representation is in part dependent on shape protrusions, how does it cope with the essentially infinite variation in potential protrusions across ‘shape space’?

One possibility is that instead of representing the shape as a whole, the shape is in fact first processed based upon its constituent parts. We are not suggesting that a shape is broken down into said parts, akin to Biederman’s geons (i.e. shape

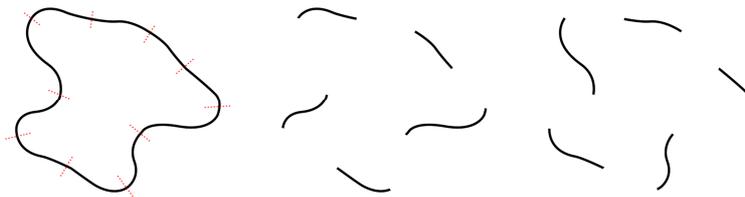
primitives such as a cylinder) from the recognition-by-components theory (Biederman, 1987). It is likely that identifying such geons would be a major computational challenge in its own right. Instead, perhaps there are filters that identify slightly convergent (or possibly closed parallel) lines in retinotopic space. These could then be used to define something akin to a shape skeleton; vectors bisecting the convergent lines, moving away from the end point. For a given shape, these vectors should themselves be largely convergent, and so sets of these convergent vectors could be taken as cohesive shape skeletons (Figure 8.1). Recognition could occur through template matching based upon characteristic skeletons. This proposed part-based representation could prove useful as greater weight could be given to more stable branches in the skeleton; i.e. the length and angle each branch forms with respect to the rest of the shape skeleton could be represented both with a mean plus estimate of variance (tolerance).



**Figure 8.1.** Hypothetical protrusion detection. We show an example shape (taken from the study outlined in Chapter 7) rendered with points of maximal curvature only (A) or the converse ‘not curves’ condition (B). The full shape is inset below in blue. Using ‘not curves’ as an example, we suggest a mechanism that detects pairs of converging lines (C). These can then be extrapolated to form protrusions, each represented with height and breadth. Vectors originating from the endpoints of these protrusions (D) should form a rough shape skeleton for recognition purposes.

To tie this back to our results, in Vernon et al. (2016) (Chapter 5), we were surprised that despite its retinotopic organisation LO-2 showed the same more ‘abstract’ representation (Shape-complexity) as LO, with no hints of the more spatially-based representation (Shape-profile) that dominated in earlier visual cortex. Furthermore, in our study comparing (amongst other conditions) curves versus ‘not curves’, we noted that LO-2 did not preferentially respond to a full shape outline despite LO and then pFs doing so; a hint of an LO-2/LO dissociation.

Whilst speculative, it is possible that LO-2's role is in fact to detect shape protrusions based upon convergent line pairs; essentially identifying a shape by its parts. LO-2's larger receptive fields (Dumoulin and Wandell, 2008; Amano et al., 2009) would allow it to identify such pairs across potentially disparate regions of the visual field, and its retinotopic organisation would prove essential to encode the orientation and location of said line pairs. These pairs could then be grouped into cohesive shapes later in LO, explaining the emerging 'full outline' sensitivity observed in this region. This could potentially be tested by again deleting 50% of the contour, but this time in alternating segments based around points of maximal curvature (Figure 8.2). This should largely remove convergent any lines, and we would therefore expect diminished activity in shape selective Lateral Occipital Cortex. Whilst it could be argued that collinearity is a confound here, we note that the original 'curves' condition (Figure 8.1A) had greater collinearity than 'not curves' (Figure 8.1B), yet still resulted in lower overall activity.



**Figure 8.2.** Deleting convergent lines. By deleting line segments adjacent to points of maximal curvature, we can largely remove any convergent lines in the object image. This makes the original shape much harder to detect, demonstrating the importance of such information.

Such a protrusion-based shape recognition mechanism could explain a number of observed results, such as the detection of illusory contours even when bounded by rounded Kanizsa-style 'pacman' shapes (Stanley and Rubin, 2003), or shape detection when parts are partially occluded (e.g. Lerner et al., 2002), providing some protrusions are still visible. It may also explain why a shape surface is preferred over a similarly shaped hole (Vinberg and Grill-Spector, 2008). If a shape is only cut into a surface as a hole, then the convergent lines will reflect the concavities in that shape, as such, the vectors would generally extend away from the shape's centroid

perhaps making recognition more difficult. Similarly, this would explain why LO generally prefers convexities over concavities (Haushofer et al., 2008a).

## 8.5 Macaque V4

An outstanding question is how the proposed mechanism relates to (or perhaps differs from) the curvature representations identified in Macaque V4? In seminal, early work, a two-dimensional shape tuning was found through direct cell recordings in awake Macaque monkeys, identifying a curvature by angular position shape space (Pasupathy and Connor, 1999, 2001, 2002; Connor et al., 2007). For example, there may be a cell tuned to acute curvature in the upper-right part of a shape. Later work suggests that this curvature tuning is largely invariant to size (El-Shamayleh and Pasupathy, 2016). Accounting for stimulus scale is important, as when an object increases in size the curvature on its boundary will become less acute and more broad; Macaque V4 appears to account for this (although some *curvature* specific responses were identified).

We note that generally the above studies used shapes with a single protrusion, this protrusion was then varied to manipulate curvature. Naturally, this means that curvature and convergent lines are confounded; tuning for acute versus broader curvature could instead reflect tuning to degree of line convergence. Pasupathy and Connor (2001) did also find a tuning bias towards acute convex curvature. Whilst this could reflect the limited set of neurons sampled, it would also be the expected result if neurons are tuned to convergent lines (broader curvature would reduce convergence).

A curvature by angular position dimensionality also leaves many open questions (although admittedly, these hold for all proposed shape representations). A measure of eccentricity (whether in retinotopic or ‘shape-centric’ coordinates) would surely also be needed, as any shape that wraps around upon itself will have multiple curvature values for a given angular position (a simple example is a doughnut shape, which has both convex and concave curvature at every angular position). When moving to 3D objects, a measure of elevation is also likely necessitated. Curvature itself may also need to be extended for 3D shapes; if we consider 2D curvature to be broadly elliptical, then 3D curvature would be ellipsoidal, requiring both major and minor axes of breadth. We have now gone from a two dimensional shape space to at

least a five dimensional space (angular position x eccentricity x elevation x major axis curvature x minor axis curvature), six dimensions if we also consider curvature height in line with our own work. Such complexity may illustrate why LO has eluded any systematic mapping thus far.

Considering LO, we note that whilst the convergent lines or protrusion-based hypothesis that we suggest and the Macaque curvature tuning evidence is largely compatible, the evidence for each is derived from different brain regions. Namely, the Macaque literature is largely drawn from V4, whereas our work focuses predominantly upon LO-1, LO-2 and LO, not human V4 (henceforth termed hV4 here for clarity). Whether hV4 is a homologue of Macaque V4 is a longstanding question, but marked differences between the two regions are observed. Most notably, Macaque V4 can be split into dorsal and ventral components, whereas hV4 appears to be a single complete hemifield map (Wandell et al., 2007). In terms of curvature tuning, early work showed that hV4 would respond not only to concentric orientations (containing curvature), but also to radial orientations that lacked curvature (Wilkinson et al., 2000). Indeed hV4 appears to prefer concentricity over curvature more generally (Dumoulin and Hess, 2007), so the evidence for curvature-specific tunings in hV4 is limited.

Our own work has shown variable results for hV4. When comparing shapes of differing complexity levels, hV4 actually showed a different response profile across standard and scrambled shapes. For standard (or recognisable) stimuli, hV4 showed a moderate retinotopic response profile, whereas neither dimension (shape-profile nor shape-complexity) really captured hV4's response profile for scrambled stimuli.

Our later follow-up with radial frequency patterns addressed this ambiguity to some extent. We found that hV4 only showed correspondence to 'lobe curvature', in contrast to LO-2 and LO which had links to lobe prominence as well as lobe curvature. As such, it is possible our scrambling process inadvertently caused differing changes along the prominence/curvature dimensions, thus eliciting differing responses in hV4. Either way, 'lobe curvature' tuning makes some sense for hV4, as this should directly correspond to concentricity; a property we know hV4 is sensitive to (Wilkinson et al., 2000; Dumoulin and Hess, 2007). That is, maintaining the height of a protrusion in a shape whilst manipulating lobe curvature

should make the protrusion change in breadth, and as breadth increases the shape will become more concentric. One surprising result is that in our experiment comparing shape boundary fragments (curves, ‘not curves’ etc.; Chapter 7), we found that for hV4 almost every condition elicited greater activation than ‘curves’. This certainly argues against hV4 as a curvature-specific region. It also raises interesting questions about concentricity, as the ‘curves’ condition contained the most concentric orientation information. Generally, we do not have enough corroborating evidence to draw firm conclusions regarding the role of hV4, indeed there is relatively little work on hV4 in the literature as a whole (Peirce, 2015). Further research is clearly needed to disentangle the roles of hV4, plus lateral occipital areas LO-1, LO-2 and LO in curvature and shape processing.

## 8.6 Concluding remarks

The work outlined in this thesis has gone some way towards outlining the relationship between retinotopic visual areas LO-1 and LO-2, plus object-selective LO. We believe there is considerable evidence for an intrinsic link between LO-2 and LO, most notably the number of instances in which we find commonality between neural representations across these two regions. Evidence regarding the role of LO-1 is mixed. We certainly have cause to believe that it is playing some independent functional role, not only from previous work (Silson et al., 2013), but also our own finding that LO-1 is the sole region to show a representation linked to radial frequency information. Despite this, LO-1 has often been found to share some aspects of the neural representations identified in later LO-2 and LO ROIs. Due to this, we would argue that both LO-1 and LO-2 represent a transitional point, where neural representations switch from a strict retinotopic basis towards something more abstract.

It is the nature of that more abstract representation that the latter half of this thesis has focused upon. Our current working hypothesis is that the representation is based upon shape protrusions. Specifically, we suggest that it is not curvature per se, but the convergent lines comorbid with curvature that shape-processing regions (perhaps LO-2) may be attuned to. Such convergent lines should indicate some degree of structure, and through extrapolation, crude shape skeletons could be generated which might form a basis for recognition. There is a large amount of work needed to further this theory. First and foremost, the ‘not curves’ versus ‘curves’ result needs to be

validated in a design better suited for probing neural representations, as that finding underpins our conclusions to a certain extent. We also need to consider how the proposed representation would extend more generally to alternative stimuli, such as 3D objects. A true account of shape tuning in LO will need considerable flexibility to capture the essentially infinite range of shape or object stimuli that we can encounter.

## References

- Amano K, Wandell BA, Dumoulin SO (2009) Visual field maps, population receptive field sizes, and visual field coverage in the human MT+ complex. *J Neurophysiol* 102:2704–2718.
- Andresen DR, Vinberg J, Grill-Spector K (2009) The representation of object viewpoint in human visual cortex. *Neuroimage* 45:522–536.
- Appelle S (1972) Perception and discrimination as a function of stimulus orientation: the “oblique effect” in man and animals. *Psychol Bull* 78:266–278.
- Arcaro MJ, McMains SA, Singer BD, Kastner S (2009) Retinotopic Organization of Human Ventral Visual Cortex. *J Neurosci* 29:10638–10652.
- Bar M, Tootell RB, Schacter DL, Greve DN, Fischl B, Mendola JD, Rosen BR, Dale AM (2001) Cortical mechanisms specific to explicit visual object recognition. *Neuron* 29:529–535.
- Baseler HA, Gouws AD, Haak K V, Racey C, Crossland MD, Tufail A, Rubin GS, Cornelissen FW, Morland AB (2011) Large-scale remapping of visual cortex is absent in adult humans with macular degeneration. *Nat Neurosci* 14:649–655.
- Bertamini M, Wagemans J (2013) Processing convexity and concavity along a 2-D contour: figure-ground, structural shape, and attention. *Psychon Bull Rev* 20:191–207.
- Biederman I (1987) Recognition-by-components: A theory of human image understanding. *Psychol Rev* 94:115–117.
- Blum H (1973) Biological shape and visual science (part I). *J Theor Biol* 38:205–287.
- Brainard DH (1997) The Psychophysics Toolbox. *Spat Vis* 10:433–436.
- Bressler DW, Silver MA (2010) Spatial attention improves reliability of fMRI retinotopic mapping signals in occipital and parietal cortex. *Neuroimage* 53:526–533.
- Brewer AA, Liu J, Wade AR, Wandell BA (2005) Visual field maps and stimulus

- selectivity in human ventral occipital cortex. *Nat Neurosci* 8:1102–1109.
- Brouwer GJ, Heeger DJ (2009) Decoding and reconstructing color from responses in human visual cortex. *J Neurosci* 29:13992–14003.
- Busing F, Commandeur JJF, Heiser WJ, Bandilla W, Faulbaum F (1997) PROXSCAL: A multidimensional scaling program for individual differences scaling with constraints. *Softstat* 97:67–74.
- Carlson ET, Rasquinha RJ, Zhang K, Connor CE (2011) A sparse object coding scheme in area V4. *Curr Biol* 21:288–293.
- Carroll JD, Chang J-J (1970) Analysis of individual differences in multidimensional scaling via an n-way generalization of “Eckart-Young” decomposition. *Psychometrika* 35:283–319.
- Connor CE, Brincat SL, Pasupathy A (2007) Transformation of shape information in the ventral pathway. *Curr Opin Neurobiol* 17:140–147.
- De Winter J, Wagemans J (2004) Contour-based object identification and segmentation: stimuli, norms and data, and software tools. *Behav Res Methods Instrum Comput* 36:604–624.
- Downing PE, Jiang Y, Shuman M, Kanwisher N (2001) A cortical area selective for visual processing of the human body. *Science* 293:2470–2473.
- Drucker DM, Aguirre GK (2009) Different spatial scales of shape similarity representation in lateral and ventral LOC. *Cereb Cortex* 19:2269–2280.
- Dumoulin SO, Hess RF (2007) Cortical specialization for concentric shape processing. *Vision Res* 47:1608–1613.
- Dumoulin SO, Wandell BA (2008) Population receptive field estimates in human visual cortex. *Neuroimage* 39:647–660.
- Duncan KJ, Devlin JT (2011) Improving the reliability of functional localizers. *Neuroimage* 57:1022–1030.
- Duncan RO, Boynton GM (2003) Cortical magnification within human primary visual cortex correlates with acuity thresholds. *Neuron* 38:659–671.

- Eger E, Ashburner J, Haynes J-D, Dolan RJ, Rees G (2008) fMRI activity patterns in human LOC carry information about object exemplars within category. *J Cogn Neurosci* 20:356–370.
- Eklund A, Nichols TE, Knutsson H (2016) Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proc Natl Acad Sci* 113:7900–7905.
- El-Shamayleh Y, Pasupathy A (2016) Contour Curvature As an Invariant Code for Objects in Visual Area V4. *J Neurosci* 36:5532–5543.
- Engel SA, Rumelhart DE, Wandell BA, Lee AT, Glover GH, Chichilnisky EJ, Shadlen MN (1994) fMRI of human visual cortex. *Nature* 369:525.
- Epstein R, Kanwisher N (1998) A cortical representation of the local visual environment. *Nature* 392:598–601.
- Etzel JA, Zacks JM, Braver TS (2013) Searchlight analysis: Promise, pitfalls, and potential. *Neuroimage* 78:261–269.
- Fang F, He S (2005) Cortical responses to invisible objects in the human dorsal and ventral pathways. *Nat Neurosci* 8:1380–1385.
- Fang F, Kersten D, Murray SO (2008) Perceptual grouping and inverse fMRI activity patterns in human visual cortex. *J Vis* 8:2–9.
- Field DJ, Hayes A, Hess RF (1993) Contour integration by the human visual system: Evidence for a local “association field.” *Vision Res* 33:173–193.
- Firestone C, Scholl BJ (2014) “Please tap the shape, anywhere you like”: Shape skeletons in human vision revealed by an exceedingly simple measure. *Psychol Sci* 25:377–386.
- Friston KJ, Rotshtein P, Geng JJ, Sterzer P, Henson RN (2006) A critique of functional localisers. *Neuroimage* 30:1077–1087.
- Glover GH (1999) Deconvolution of impulse response in event-related BOLD fMRI. *Neuroimage* 9:416–429.
- Goodale MA, Milner AD (1992) Separate visual pathways for perception and action.

Trends Neurosci 15:20–25.

Gouws AD, Alvarez I, Watson DM, Uesaki M, Rodgers J, Morland AB (2014) On the Role of Suppression in Spatial Attention: Evidence from Negative BOLD in Human Subcortical and Cortical Structures. *J Neurosci* 34:10347–10360.

Grill-Spector K, Kourtzi Z, Kanwisher N (2001) The lateral occipital complex and its role in object recognition. *Vision Res* 41:1409–1422.

Grill-Spector K, Kushnir T, Edelman S, Avidan G, Itzhak Y, Malach R (1999) Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron* 24:187–203.

Grill-Spector K, Kushnir T, Edelman S, Itzhak Y, Malach R (1998a) Cue-invariant activation in object-related areas of the human occipital lobe. *Neuron* 21:191–202.

Grill-Spector K, Kushnir T, Hendler T, Edelman S, Itzhak Y, Malach R (1998b) A sequence of object-processing stages revealed by fMRI in the human occipital lobe. *Hum Brain Mapp* 6:316–328.

Grill-Spector K, Kushnir T, Hendler T, Malach R (2000) The dynamics of object-selective activation correlate with recognition performance in humans. *Nat Neurosci* 3:837–843.

Grill-Spector K, Weiner KS (2014) The functional architecture of the ventral temporal cortex and its role in categorization. *Nat Rev Neurosci* 15:536–548.

Habak C, Wilkinson F, Zakher B, Wilson HR (2004) Curvature population coding for complex shapes in human vision. *Vision Res* 44:2815–2823.

Hansen KA, David S V., Gallant JL (2004) Parametric reverse correlation reveals spatial linearity of retinotopic human V1 BOLD response. *Neuroimage* 23:233–241.

Harmon LD, Julesz B (1973) Masking in visual recognition: effects of two dimensional filtered noise. *Science* (80- ) 180:1194–1197.

Hasson U, Levy I, Behrmann M, Hendler T, Malach R (2002) Eccentricity bias as an

- organizing principle for human high-order object areas. *Neuron* 34:479–490.
- Haushofer J, Baker CI, Livingstone MS, Kanwisher N (2008a) Privileged coding of convex shapes in human object-selective cortex. *J Neurophysiol* 100:753–762.
- Haushofer J, Livingstone MS, Kanwisher N (2008b) Multivariate patterns in object-selective cortex dissociate perceptual and physical shape similarity. *PLoS Biol* 6:e187.
- Haxby J V, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* (80- ) 293:2425–2430.
- Haynes J-D, Rees G (2005) Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nat Neurosci* 8:686–691.
- Hayworth KJ, Biederman I (2006) Neural evidence for intermediate representations in object recognition. *Vision Res* 46:4024–4031.
- Hemond CC, Kanwisher N, Op de Beeck HP (2007) A preference for contralateral stimuli in human object- and face-selective cortex. Ferrari P, ed. *PLoS One* 2:e574.
- Hess RF, Hayes A, Field DJ (2003) Contour integration and cortical processing. *J Physiol Paris* 97:105–119.
- Holmes G (1918) Disturbances of Vision by Cerebral Lesions. *Br J Ophthalmol* 2:353–384.
- Horn JL (1965) A rationale and test for the number of factors in factor analysis. *Psychometrika* 30:179–185.
- Hubel DH, Wiesel TN (1959) Receptive fields of single neurones in the cat's striate cortex. *J Physiol* 148:574–591.
- Huetzel SA, Song AW, McCarthy G (2009) MR Contrast Mechanisms and Pulse Sequences. In: *Functional Magnetic Resonance Imaging*, 2nd Ed., pp 123. Sunderland Mass.: Sinauer Associates.
- Hupé J-M, Bordier C, Dojat M (2012) A BOLD signature of eyeblinks in the visual

cortex. *Neuroimage* 61:149–161.

James TW, Culham J, Humphrey GK, Milner AD, Goodale MA (2003) Ventral occipital lesions impair object recognition but not object-directed grasping: an fMRI study. *Brain* 126:2463–2475.

Jenkinson M, Bannister P, Brady M, Smith SM (2002) Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* 17:825–841.

Jenkinson M, Smith SM (2001) A global optimisation method for robust affine registration of brain images. *Med Image Anal* 5:143–156.

Jones HE, Grieve KL, Wang W, Sillito AM (2001) Surround Suppression in Primate V1. *J Neurophysiol* 86:2011–2028.

Julian JB, Fedorenko E, Webster J, Kanwisher N (2012) An algorithmic method for functionally defining regions of interest in the ventral visual pathway. *Neuroimage* 60:2357–2364.

Kaiser HF (1960) The Application of Electronic Computers to Factor Analysis. *Educ Psychol Meas* 20:141–151.

Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci* 17:4302–4311.

Kanwisher N, Wojciulik E (2000) Visual attention: insights from brain imaging. *Nat Rev Neurosci* 1:91–100.

Kayaert G, Biederman I, Op de Beeck HP, Vogels R (2005) Tuning for shape dimensions in macaque inferior temporal cortex. *Eur J Neurosci* 22:212–224.

Kayaert G, Wagemans J, Vogels R (2011) Encoding of complexity, shape, and curvature by macaque infero-temporal neurons. *Front Syst Neurosci* 5:51.

Kim JG, Biederman I, Lescroart MD, Hayworth KJ (2009) Adaptation to objects in the lateral occipital complex (LOC): shape or semantics? *Vision Res* 49:2297–2305.

- Kingdom FAA, Whittle P (1996) Contrast discrimination at high contrasts reveals the influence of local light adaptation on contrast processing. *Vision Res* 36:817–829.
- Kok P, Jehee JFM, de Lange FP (2012) Less Is More: Expectation Sharpens Representations in the Primary Visual Cortex. *Neuron* 75:265–270.
- Kolster H, Peeters R, Orban GA (2010) The retinotopic organization of the human middle temporal area MT/V5 and its cortical neighbors. *J Neurosci* 30:9801–9820.
- Kourtzi Z, Erb M, Grodd W, Bühlhoff HH (2003) Representation of the perceived 3-D object shape in the human lateral occipital complex. *Cereb cortex* 13:911–920.
- Kourtzi Z, Kanwisher N (2000) Cortical regions involved in perceiving object shape. *J Neurosci* 20:3310–3318.
- Kourtzi Z, Kanwisher N (2001) Representation of perceived object shape by the human lateral occipital complex. *Science* (80- ) 293:1506–1509.
- Kravitz DJ, Kriegeskorte N, Baker CI (2010) High-level visual object representations are constrained by position. *Cereb Cortex* 20:2916–2925.
- Kriegeskorte N, Mur M, Bandettini P a. (2008) Representational similarity analysis - connecting the branches of systems neuroscience. *Front Syst Neurosci* 2:4.
- Kriegeskorte N, Simmons WK, Bellgowan PSF, Baker CI (2009) Circular analysis in systems neuroscience: the dangers of double dipping. *Nat Neurosci* 12:535–540.
- Kurki I, Saarinen J, Hyvärinen A (2014) Investigating shape perception by classification images. *J Vis* 14:1–19.
- Larsson J, Heeger DJ (2006) Two retinotopic visual areas in human lateral occipital cortex. *J Neurosci* 26:13128–13142.
- Larsson J, Heeger DJ, Landy MS (2010) Orientation selectivity of motion-boundary responses in human visual cortex. *J Neurophysiol* 104:2940–2950.

- Larsson J, Landy MS, Heeger DJ (2006) Orientation-selective adaptation to first- and second-order patterns in human visual cortex. *J Neurophysiol* 95:862–881.
- Lerner Y, Hendler T, Ben-Bashat D, Harel M, Malach R (2001) A hierarchical axis of object processing stages in the human visual cortex. *Cereb cortex* 11:287–297.
- Lerner Y, Hendler T, Malach R (2002) Object-completion effects in the human lateral occipital complex. *Cereb cortex* 12:163–177.
- Levy I, Hasson U, Avidan G, Hendler T, Malach R (2001) Center-periphery organization of human object areas. *Nat Neurosci* 4:533–539.
- Loffler G (2008) Perception of contours and shapes: low and intermediate stage mechanisms. *Vision Res* 48:2106–2127.
- Loffler G, Wilson HR, Wilkinson F (2003) Local and global contributions to shape discrimination. *Vision Res* 43:519–530.
- Logothetis NK, Wandell BA (2004) Interpreting the BOLD signal. *Annu Rev Physiol* 66:735–769.
- Malach R, Levy I, Hasson U (2002) The topography of high-order human object areas. *Trends Cogn Sci* 6:176–184.
- Malach R, Reppas JB, Benson RR, Kwong KK, Jiang H, Kennedy WA, Ledden PJ, Brady TJ, Rosen BR, Tootell RB (1995) Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proc Natl Acad Sci U S A* 92:8135–8139.
- Mendola JD, Dale AM, Fischl B, Liu AK, Tootell RB (1999) The representation of illusory and real contours in human cortical visual areas revealed by functional magnetic resonance imaging. *J Neurosci* 19:8560–8572.
- Montaser-Kouhsari L, Landy MS, Heeger DJ, Larsson J (2007) Orientation-selective adaptation to illusory contours in human visual cortex. *J Neurosci* 27:2186–2195.
- Morgan AT, Petro LS, Muckli L (2016) Cortical feedback to V1 and V2 contains

- unique information about high-level scene structure. *BioRxiv*:41186.
- Mumford J (2007) A guide to calculating percent change with featquery.
- Murray MM, Wylie GR, Higgins BA, Javitt DC, Schroeder CE, Foxe JJ (2002a) The spatiotemporal dynamics of illusory contour processing: combined high-density electrical mapping, source analysis, and functional magnetic resonance imaging. *J Neurosci* 22:5055–5073.
- Murray SO, Kersten D, Olshausen B a, Schrater P, Woods DL (2002b) Shape perception reduces activity in human primary visual cortex. *Proc Natl Acad Sci U S A* 99:15164–15169.
- Niemeier M, Goltz HC, Kuchinad A, Tweed DB, Vilis T (2005) A contralateral preference in the lateral occipital area: Sensory and attentional mechanisms. *Cereb Cortex* 15:325–331.
- Ogawa S, Lee TM, Kay AR, Tank DW (1990) Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proc Natl Acad Sci U S A* 87:9868–9872.
- Oliva A, Torralba A (2001) Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope. *Int J Comput Vis* 42:145–175.
- Op de Beeck HP, Baker CI, DiCarlo JJ, Kanwisher N (2006) Discrimination training alters object representations in human extrastriate cortex. *J Neurosci* 26:13025–13036.
- Op de Beeck HP, Haushofer J, Kanwisher N (2008a) Interpreting fMRI data: maps, modules and dimensions. *Nat Rev Neurosci* 9:123–135.
- Op de Beeck HP, Torfs K, Wagemans J (2008b) Perceived shape similarity among unfamiliar objects and the organization of the human object vision pathway. *J Neurosci* 28:10111–10123.
- Panis S, De Winter J, Vandekerckhove J, Wagemans J (2008) Identification of everyday objects on the basis of fragmented outline versions. *Perception* 37:271–289.

- Pasupathy A, Connor CE (1999) Responses to contour features in macaque area V4. *J Neurophysiol* 82:2490–2502.
- Pasupathy A, Connor CE (2001) Shape representation in area V4: position-specific tuning for boundary conformation. *J Neurophysiol* 86:2505–2519.
- Pasupathy A, Connor CE (2002) Population coding of shape in area V4. *Nat Neurosci* 5:1332–1338.
- Peirce JW (2015) Understanding mid-level representations in visual processing. *J Vis* 15:5.
- Pelli DG (1997) The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat Vis* 10:437–442.
- Pitcher D, Charles L, Devlin JT, Walsh V, Duchaine B (2009) Triple Dissociation of Faces, Bodies, and Objects in Extrastriate Cortex. *Curr Biol* 19:319–324.
- Poirier FJAM, Wilson HR (2007) Object perception and masking: Contributions of sides and convexities. *Vision Res* 47:3001–3011.
- Quiroga RQ, Reddy L, Kreiman G, Koch C, Fried I (2005) Invariant visual representation by single neurons in the human brain. *Nature* 435:1102–1107.
- Regan BC, Julliot C, Simmen B, Viénot F, Charles-Dominique P, Mollon JD (2001) Fruits, foliage and the evolution of primate colour vision. *Philos Trans R Soc Lond B Biol Sci* 356:229–283.
- Rensink RA, Baldrige G (2010) The Perception of Correlation in Scatterplots. *Comput Graph Forum* 29:1203–1210.
- Riddoch G (1917) Dissociation of visual perceptions due to occipital injuries, with especial reference to appreciation of movement. *Brain* 40:15–57.
- Ritzl A, Marshall JC, Weiss PH, Zafiris O, Shah NJ, Zilles K, Fink GR (2003) Functional anatomy and differential time courses of neural processing for explicit, inferred, and illusory contours. An event-related fMRI study. *Neuroimage* 19:1567–1577.
- Salmela VR, Henriksson L, Vanni S (2016) Radial Frequency Analysis of Contour

- Shapes in the Visual Cortex Einhäuser W, ed. PLoS Comput Biol 12:e1004719.
- Sapountzis P, Schluppeck D, Bowtell R, Peirce JW (2010) A comparison of fMRI adaptation and multivariate pattern classification analysis in visual cortex. *Neuroimage* 49:1632–1640.
- Sawamura H, Georgieva S, Vogels R, Vanduffel W, Orban GA (2005) Using functional magnetic resonance imaging to assess adaptation and size invariance of shape processing by humans and monkeys. *J Neurosci* 25:4294–4306.
- Saxe R, Brett M, Kanwisher N (2006) Divide and conquer: a defense of functional localizers. *Neuroimage* 30:1088-96-9.
- Sayres RA, Grill-Spector K (2008) Relating retinotopic and object-selective responses in human lateral occipital cortex. *J Neurophysiol* 100:249–267.
- Schwarzlose RF, Swisher JD, Dang S, Kanwisher N (2008) The distribution of category and location information across object-selective regions in human visual cortex. *Proc Natl Acad Sci* 105:4447–4452.
- Self MW, Zeki S (2005) The integration of colour and motion by the human visual brain. *Cereb cortex* 15:1270–1279.
- Sereno MI, Dale AM, Reppas JB, Kwong KK, Belliveau JW, Brady TJ, Rosen BR, Tootell RB (1995) Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science* (80- ) 268:889–893.
- Shpaner M, Molholm S, Forde E, Foxe JJ (2013) Disambiguating the roles of area V1 and the lateral occipital complex (LOC) in contour integration. *Neuroimage* 69:146–156.
- Silson EH, Groen IIA, Kravitz DJ, Baker CI (2016) Evaluating the correspondence between face-, scene-, and object-selectivity and retinotopic organization within lateral occipitotemporal cortex. *J Vis* 16:1–21.
- Silson EH, McKeefry DJ, Rodgers J, Gouws AD, Hymers M, Morland AB (2013) Specialized and independent processing of orientation and shape in visual field maps LO1 and LO2. *Nat Neurosci* 16:267–269.

- Smith SM (2002) Fast robust automated brain extraction. *Hum Brain Mapp* 17:143–155.
- Snodgrass JG, Vanderwart M (1980) A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *J Exp Psychol Hum Learn* 6:174–215.
- Snow JC, Pettypiece CE, McAdam TD, McLean AD, Stroman PW, Goodale MA, Culham J (2011) Bringing the real world into the fMRI scanner: repetition effects for pictures versus real objects. *Sci Rep* 1:130.
- Stanley DA, Rubin N (2003) fMRI activation in response to illusory contours and salient regions in the human lateral occipital complex. *Neuron* 37:323–331.
- Tootell RB, Mendola JD, Hadjikhani NK, Ledden PJ, Liu AK, Reppas JB, Sereno MI, Dale AM (1997) Functional analysis of V3A and related areas in human visual cortex. *J Neurosci* 17:7060–7078.
- Tootell RBH (2001) Where is “Dorsal V4” in Human Visual Cortex? Retinotopic, Topographic and Functional Evidence. *Cereb Cortex* 11:298–311.
- Vernon RJW, Gouws AD, Lawrence SJD, Wade AR, Morland AB (2016) Multivariate Patterns in the Human Object-Processing Pathway Reveal a Shift from Retinotopic to Shape Curvature Representations in Lateral Occipital Areas, LO-1 and LO-2. *J Neurosci* 36:5763–5774.
- Vinberg J, Grill-Spector K (2008) Representation of shapes, edges, and surfaces across multiple cues in the human visual cortex. *J Neurophysiol* 99:1380–1393.
- Vuilleumier P, Henson RN, Driver J, Dolan RJ (2002) Multiple levels of visual object constancy revealed by event-related fMRI of repetition priming. *Nat Neurosci* 5:491–499.
- Wandell BA, Dumoulin SO, Brewer AA (2007) Visual field maps in human cortex. *Neuron* 56:366–383.
- Weber M, Thompson-Schill SL, Osherson D, Haxby J V, Parsons L (2009) Predicting judged similarity of natural categories from their neural representations. *Neuropsychologia* 47:859–868.

- Wertheimer M (1923) Untersuchungen zur Lehre von der Gestalt, II. *Psychol Forsch* 4:301–350.
- Wilkinson F, James TW, Wilson HR, Gati JS, Menon RS, Goodale MA (2000) An fMRI study of the selective activation of human extrastriate form vision areas by radial and concentric gratings. *Curr Biol* 10:1455–1458.
- Wilkinson F, Wilson HR, Habak C (1998) Detection and recognition of radial frequency patterns. *Vision Res* 38:3555–3568.
- Williams MA, Dang S, Kanwisher N (2007) Only some spatial patterns of fMRI response are read out in task performance. *Nat Neurosci* 10:685–686.
- Worsley KJ (2001) Statistical Analysis of Activation Images. In: *Functional MRI: An Introduction to Methods* (Jezzard P, Matthews PM, Smith SM, eds), pp 251–270. Oxford: Oxford University Press.
- Yue X, Pourladian IS, Tootell RBH, Ungerleider LG (2014) Curvature-processing network in macaque visual cortex. *Proc Natl Acad Sci U S A* 111:E3467-75.
- Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, Gerig G (2006) User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *Neuroimage* 31:1116–1128.
- Zahn CT, Roskies RZ (1972) Fourier Descriptors for Plane Closed Curves. *IEEE Trans Comput C-21*:269–281.
- Zarahn E, Aguirre GK, D’Esposito M (1997) Empirical Analyses of BOLD fMRI Statistics. *Neuroimage* 5:179–197.
- Zhang Y, Brady M, Smith S (2001) Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans Med Imaging* 20:45–57.