

SOLID SHAPE PROCESSING IN AREA V4

by

Ramanujan Srinath

A dissertation submitted to Johns Hopkins University in conformity
with the requirements for the degree of Doctor of Philosophy.

Baltimore, Maryland

August 2019

© 2019 Ramanujan Srinath

All rights reserved

Abstract

Coding transformations in ventral visual cortex convert image information into knowledge about object structure and identity. Neurons in area V4, an intermediate ventral pathway stage in the macaque brain, are known to represent 2D contour, binocular disparity, and 3D orientation. In the experiments outlined in this thesis, I address fundamental questions about the relationship between 2D image signals and solid shape perception in the ventral visual stream.

In chapter 1, I demonstrate that a substantial fraction of V4 neurons are more responsive to solid 3D than to planar 2D shape. I constructed simple solid shape forms around medial axis skeletons having randomized connectivity, orientation, curvature, and surface structure, and I leveraged realistic shading cues to render them as solid objects. Using micro-electrode recording in awake, fixating monkeys, I analyzed the tuning of V4 neurons to these solid shapes and their planar silhouettes and found that most V4 neurons ($\sim 70\%$) prefer solid shape fragments. I then parameterized stimulus shapes in terms of local geometric measures of contours, surfaces, and medial axes, producing a point cloud description of each stimulus in a multi-dimensional geometric space. Analysis of this space suggests that solid shape models more generally account for V4 responses, while slices through these models might explain responses to planar shapes.

In chapter 2, I describe two-photon imaging experiments in V4 in anesthetized animals using OGB to measure neuronal responses. Again, a substantial fraction of neurons responded more strongly to solid shapes than to their planar silhouettes. I observed strong local clustering of solid-/planar-preferring neurons in separate cortical patches, and neighboring patches were most responsive to congruent solid and planar shapes. This suggests that derivation of solid shape from image information is a constraint on micro-organization in area V4.

Previous studies using gratings, complex planar shapes, and naturalistic images fail to address whether shape processing is inherently 3D, or whether planar and solid shape are processed in parallel. By rigorously studying V4 with solid shape stimuli, we can postulate general principles of coding transformations in the ventral visual pathway.

Readers:

Charles E. Connor, Ph. D. (advisor, first reader)

Kristina J. Nielsen, Ph. D. (advisor, second reader)

Thesis committee:

James J. Knierim, Ph. D. (chair)

Shreesh Mysore, Ph. D.

Rudiger von der Heydt, Ph. D.

Acknowledgements

Acknowledgements are meaningless without some context around what the achievement signifies. Although I have failed at similar junctures to properly balance self-justification and modesty, here, I will err on the side of self-effacing; this thesis is truly dedicated to everyone named below.

For the first twenty years of my life, I took great pride in being a generalist – the jack of a few trades. I could do a neat card trick, I could make a mean omelette, I could make a 4-bit up/down synchronous counter with just NAND gates. Now, while *I* cannot justify calling myself a specialist, this degree deigns to do just that – literally, it means ‘a specialist in thinking about neuroscience’. In a way, this experience has filled a long, *long* standing void in my heart; the desire to continue to be the jack of a few trades, but at least be the *certified* master of one.

For this honor, I would like to thank my advisors – Drs. Kristina Nielsen and Ed Connor – for believing that, given an opportunity, someone with a vision (*ba-dum-tish*), however ill-conceived and impractical, could learn to add value to a field of research. Their incessant efforts towards honing my technical capabilities, and bottomless patience for my ‘what-if’s and ‘perhaps-I-could’s was invaluable my years in their labs. If there is an astute reader familiar with Ed and/or Kristina, they will be able to see, in parts, traces, and, in parts, gobs, of their genius and rigor throughout this document. Working with them, I have learnt invaluable and concrete lessons on scientific thinking, rigor, conduct, experimental design, and mathematical analyses, and more fuzzy lessons on how to be a functional, well-adjusted, and positive individual, contributing actively to neuroscientific research.

My thesis committee, Drs. James Knierim, Shreesh Mysore, and Rudiger von der Heydt, have been a constant source of feedback, support and guidance. This thesis could not have been realized in this form without them. A huge thanks to my co-workers, Augusto Lempel, Erika Dunn-Weiss, and Alexandriya Emonds, and Andrew Cheng for their patience with my rambling journey through scientific arguments and countless hours of commiseration through my travails in graduate school. Their viewpoints have shaped this thesis for the better in immeasurable ways. I would like to thank the MBI staff, Ofelia Garalde, Justin Killibrew, Bill Nash, Bill Quinlan, for technical assistance for my experiments and analyses, and Rita Ragan, Beth Wood-Roig, Susan Soohoo, and Eric Potter for keeping the graduate school machinery working in top condition always.

On a good day, graduate school was exhilarating. All my experiments and analyses fell into place and I discovered something shiny. On every other day, it would have been unbearable had it not been for my battalion of friends willing to put up with my whinging and whining. I owe a great debt of gratitude to Richard, Gabby, Aneesh, Chanel, Alice, and my connections to India in a foreign land, Purnima, Veena, Mithra, Akshay, Nagaraj and Vimal.

A huge force behind my being able to finish this thesis with some semblance of dignity and decorum is my girlfriend, Alexandriya. Her undying love and patience helped me suffer through the trials of experiments and analyses. She was my guide and mentor when I needed support, and my sounding board when I needed to vent. Also, half the commas in this document are courtesy of her – Alexandriya picked nits off of this document I didn't even know existed. I look forward to celebrating a new beginning, when I have some sweet, sweet post-doc money to burn.

Finally, and primarily, I have to thank my family. My mother's dogged pursuit for the well-being of her sons has had a visibly lasting effect on my physical and mental strength. My father's

everlasting determination to help me achieve my goals – goals he helped me set in the first place – and be my confidante through life, has given me the courage to believe that I could set my mind to something and perhaps even excel. My brother’s boundless love and affection has been a place of respite after many a long day, and I dedicate this thesis to him. Special thanks to my uncle for vacations on which I gained more perspective on the world than my years in grad school, and to my grandmother who has always made me feel at home at home. In twenty years, I will look back at my rapid growth as a human over the past seven years and give full credit to my family. This, along with everything else in life, would not have been possible without them.

Publications and co-authors

Srinath, Emonds, Wang, Nielsen, Connor. (2019). Volumes, not patterns, are the basic elements of object vision. (*submitted*)

Srinath, Emonds, Lempel, Dunn-Weiss, Connor, Nielsen (2019). Clustering of 3D shape information in V4. (*in preparation*)

Table of Contents

Abstract.....	ii
Acknowledgements	iv
Table of Contents	vii
List of Figures.....	x
1 Introduction	1
1.1 Cortical coding transformation hierarchy	1
1.2 Organization of primate object vision	3
1.2.1 Shape representation in early object vision.....	3
1.2.2 Shape representation in V4	5
1.3 Solid shape vision.....	8
1.3.1 Solid shape from various cues	8
1.3.2 Solid shape in V4	10
1.3.3 Solid shape processing beyond V4	12
1.4 Micro-organization of visual areas	13
1.4.1 Functional organization and maps in primary visual cortex.....	14
1.4.2 The utility of cortical maps and their mutual alignment.....	17
1.4.3 Functional maps in higher ventral stream areas.....	20
1.5 Studying solid shape representation in V4	22
2 Solid shape representation in V4	25
2.1 Introduction	25
2.1.1 Experimental approaches to studying shape coding	26

2.2	Results	28
2.2.1	2D and 3D shape information in area V4	28
2.2.2	Parameterizing shape space and visualizing shape tuning.....	32
2.2.3	Robust representation across 3D cues.....	37
2.3	Discussion	41
2.3.1	Solid shape coding unifies contour and spectral coding hypotheses	41
2.3.2	The need for and utility of solid shape processing in V4.....	44
2.4	Methods	46
2.4.1	Experimental Model and Subject Details.....	46
2.4.2	Method Details	46
2.4.3	Response-weighted Averaging Analysis	49
3	Micro-organization of solid shape representation	51
3.1	Introduction	51
3.2	Results	52
3.2.1	Neural responses to solid shapes with two-photon imaging	52
3.2.2	Organization of solid shape preference.....	59
3.2.3	Comparisons of shape preferences within and across clusters.....	66
3.3	Discussion	70
3.3.1	Relationship to known maps in V4	72
3.3.2	Implications for V4 inputs and outputs	73
3.4	Methods	74
3.4.1	Experimental Model and Subject Details.....	74
3.4.2	Method Details	74
3.4.3	Quantification and Statistical Analysis	79
4	Discussion	82

4.1	Image \Rightarrow Solid Shape \Rightarrow Identity	83
4.1.1	Deep network models of object vision.....	83
4.2	Future directions	86
4.2.1	Population coding and dynamics of solid shape representation	86
4.2.2	Local and global circuits for solid shape processing	88
4.3	Conclusion	91
5	Appendix 1: Solid shape parameterization	93
6	Appendix 2: Supplementary Figures	95
6.1	Supplementary figures for chapter 2	95
6.2	Supplementary figures for chapter 3	98
7	Bibliography	100
8	Curriculum Vitae	118

List of Figures

Figure 1.1: Hierarchy of areas in the ventral visual pathway.....	1
Figure 1.2: RF sizes, response latencies, and emergent tuning properties of ventral visual pathway areas. ...	2
Figure 1.3: Absolute and relative disparity and their contribution to depth and solid form processing.	7
Figure 1.4: Shading and stereoscopic cues for solid shape perception.	9
Figure 1.5: Functional maps in V1 and pooling hypotheses.	16
Figure 2.1: Adaptive solid shape sampling experiment.	31
Figure 2.2: Metric shape analysis.....	32
Figure 2.3: Response-weighted averaging (RWA) analysis.	33
Figure 2.4: RWA matrix for an example neuron.....	35
Figure 2.5: Examples of solid shape-preferring cells with tuning for solid shape fragments.....	36
Figure 2.6: Population analysis of SP and RWA.	37
Figure 2.7: Contrast post-hoc test.....	38
Figure 2.8: Naturalistic images post-hoc test.	39
Figure 2.9: Stereogram post-hoc test.....	40
Figure 3.1: Example two-photon imaging region 1.	53
Figure 3.2: Receptive field mapping during two-photon imaging.	54
Figure 3.3: Trial structure and stimulus set 1 used in two-photon imaging experiments.	56
Figure 3.4: Trial structure and stimulus set 2 used in two-photon imaging experiments.	57
Figure 3.5: Fluorescence time course for solid and planar shape-preferring neurons.	59
Figure 3.6: An example functional map of solid shape preference in V4.....	60
Figure 3.7: Spatial clustering analysis based on pixel-SP functional map.....	61
Figure 3.8: Spatial clustering and transition of SP along cluster boundaries.....	62
Figure 3.9: Cumulative distribution of SP in two-photon imaging experiments.	63
Figure 3.10: Spatial clustering of SP for four more imaging regions.	65

Figure 3.11: Correlational analyses for all five imaging regions.....	67
Figure 4.1: Neural networks and latent spaces.....	84
Figure 5.1: Parameterization of surface fragments and medial axial components.....	94
Figure 6.1: Examples of neurons highly selective for planar silhouette stimuli.....	95
Figure 6.2: Example of a full set of stimuli in the naturalistic images post-hoc.....	96
Figure 6.3: Example stereograms used in the stereogram post-hoc test.....	97
Figure 6.4: Pairwise correlation distributions for neurons within clusters.....	98
Figure 6.5: Schematics explaining the three types of correlational analysis in section 3.2.3.....	99

1 Introduction

We live in a world replete with objects with which we interact fluidly and adeptly. Within a fraction of a second of looking at a completely novel object, we can immediately make accurate inferences about its shape, texture, weight, stability, etc. – in essence, its *identity*. We seem to be able to visually perceive a wooden block – simple, inanimate, and static – just as quickly and effortlessly as a human body – complex, articulate, and dynamic. Underlying this ability is one of the most computationally intricate processes our brains are tasked to perform, and the set of cortical areas that accomplish this task is called the ventral visual pathway.

1.1 Cortical coding transformation hierarchy

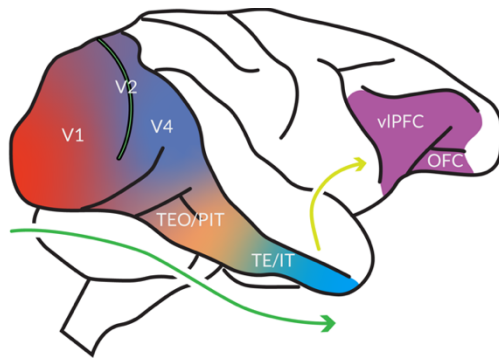


Figure 1.1: Hierarchy of areas in the ventral visual pathway.

Ventral visual pathway areas defined by retinotopy and functional properties. Areas in prefrontal cortex are marked as targets of area TE and are not a part of the visual hierarchy, per se.

The ultimate goal of the ventral visual pathway is to represent the visual world in a way that can be easily decoded by other areas. To that end, the ventral visual pathway converts dense, variable, and pixelated input from the eyes into object identity by applying a set of coding

transformations to the input. These transformations result in a

representation of the visual world that

is robust to changes in viewing conditions and compacted for storage in (and retrieval from)

memory. This compact, and explicit representation of the visual world aids the evaluation of

object value and action-associations by frontal cortex areas like pre-frontal cortex (vIPFC),

orbitofrontal cortex (OFC) etc. Retinal input is relayed to area primary visual cortex (V1) by the

lateral geniculate nucleus (LGN), where this process begins. V1 extracts information like the orientation, spatial frequency, disparity, color and direction, of small, local regions in the visual field (see figure 1.2 for a schematic of ventral visual pathway representations starting with V1). In intermediate visual areas V2 through V4, this information is sequentially transformed to represent larger and more complex object features like sharp or smooth curvature, relative disparity, complex textures etc. Similarly, the multiple stages of inferotemporal cortex (IT) represent object information in the form of complex combinations of relatively simpler features. These transformations, from V1 to IT, serve to make the visual representation more *explicit*, i.e. easier to read out by downstream areas, and more *compact*, i.e. any information can

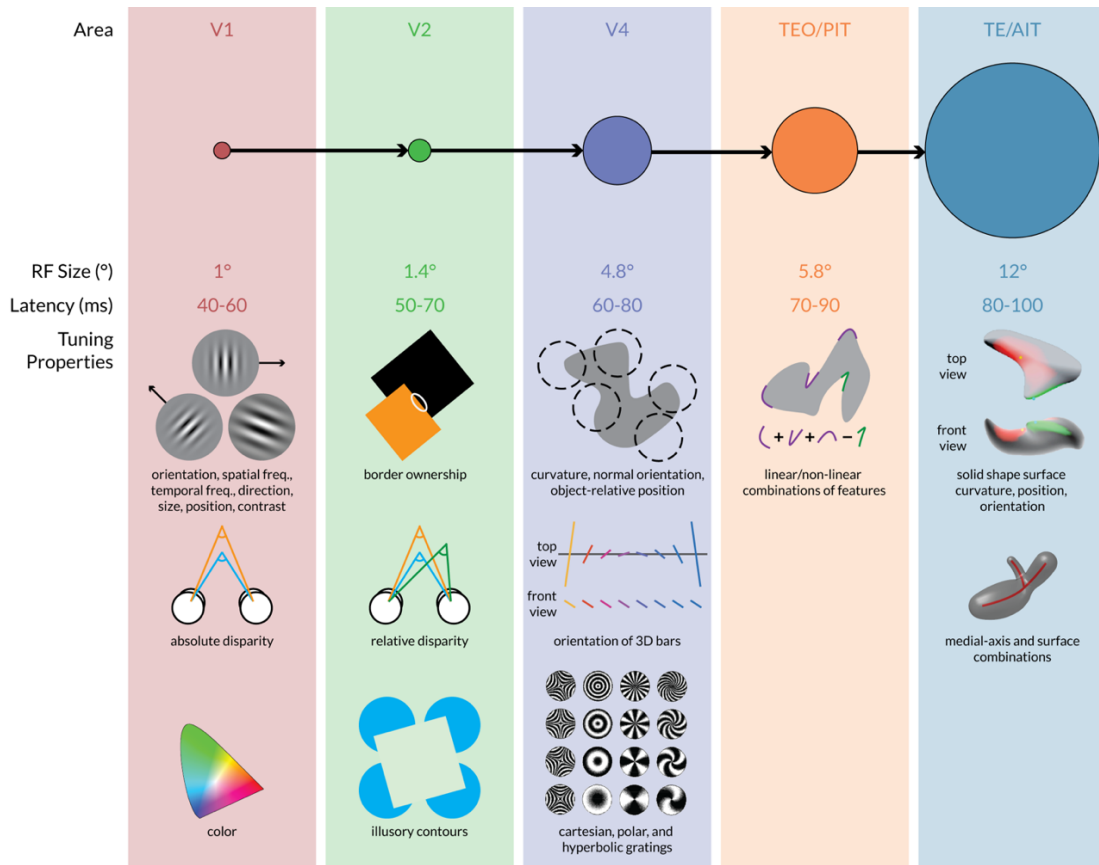


Figure 1.2: RF sizes, response latencies, and emergent tuning properties of ventral visual pathway areas.

be read out by monitoring a few neurons. (See (Felleman and Van Essen, 1991) and (Kravitz et al., 2013) for a review of the areas in the ventral visual pathway.) The neural algorithms implementing these transformations can be studied with neurophysiological techniques, like electrophysiology or optical/two-photon imaging, in monkeys. Monkeys have high visual acuity and capacity for visual shape recognition, with a ventral visual pathway that is remarkably analogous to that of humans.

The intermediate areas of the ventral pathway are a particularly interesting component of this pathway, because they implement some of the key transformations from the relatively simple features represented in their input areas into the abstract and complex representations found in the highest pathway stages. The understanding of these intermediate transformations is vital for creating models of object vision, machine vision applications, and visual prosthetics. This thesis focuses on the representation of 3D solid shape information in area V4. The following section outlines recent advances in describing the neural algorithms underlying primate object vision, with special emphasis on object shape coding in area V4. Other aspects of object identity, like material, texture, color, stability, animacy, etc., while extremely important and interesting, are either believed to be coded in higher visual areas or are beyond the scope of this thesis.

1.2 Organization of primate object vision

1.2.1 Shape representation in early object vision

Representation of the visual world starts off pixelated and dense in the primate **retina**. Retinal ganglion cells (RGCs) respond to light/dark (and/or chromatic) contrast in a small part of the visual field called the receptive field (RF). This representation is not unlike a digital photograph, where each pixel corresponds to the response of an RGC. Input from both eyes reaches both

halves of the **LGN** but remains segregated. Here, for the purposes of this thesis, relatively minor transformations occur, and the information is effectively relayed to primary visual cortex (**V1**).

V1 extracts edge and orientation information from the visual scene. Neurons in **V1** have small receptive fields at the fovea, on the order of 0.5° of visual angle, and can be approximated as spatio-temporal spectral filters tuned to orientation, spatial frequency, spatial phase, temporal frequency, contrast, color, disparity, and direction (De Valois et al., 1982; Goris et al., 2015; Hubel and Wiesel, 1962; Levitt and Lund, 1997; Movshon et al., 1978; Schiller et al., 1976; Skottun et al., 1991). This encoding is likely achieved by integrating inputs from neighboring **LGN** neurons all tuned for similar visual properties (Paik and Ringach, 2011). Contour edge detection is a major first step in compacting visual representation and aids image segmentation.

V2, a major target of neural projections from **V1**, represents combinations of simple features, like sharp and broad angles and curvatures (Hegdé and Van Essen, 2000, 2004, 2003) and illusory contours (von der Heydt et al., 1984). **V2** neurons have larger receptive fields (RFs) and are therefore amenable to the encoding of changes in contour orientation. Additionally, owing to larger RFs and targeted feedback from downstream areas, **V2** neurons are modulated by whether the contour in the RF is perceived to be a part of the foreground object or the background (Zhou et al., 2000). This is called encoding border ownership and is the basis of image segmentation by figure-ground discrimination and tuning for object-relative position in later areas. **V2** neurons also encode the depth (absolute disparity) of a stereoscopic edge (von der Heydt et al., 2000) and the depth of edges relative to other edges (relative disparity) (Cumming and Parker, 1999) possibly by integrating across binocular disparity-tuned neurons in **V1**. Relative disparity coding of edges in **V2** probably leads to the computation of the relative disparity of shapes in later areas of the ventral visual pathway (see figure 1.3).

1.2.2 Shape representation in V4

Area V4 is a mid-tier object processing region that is known to represent complex object shape properties. Neurons in V4 integrate inputs from early visual areas V1 and V2, have considerably larger RFs (Gattass et al., 1988), and exhibit tuning to a diverse set of object properties like surface texture, object form (orientation and curvature), relative disparity and depth, and color. A brief review of the literature is presented below.

Areal extent and retinotopy: Area V4 was originally distinguished from neighboring areas V1 and MT on the basis of its connections with V2 (Zeki, 1971) and on the basis of its functional selectivity for color (Zeki, 1973). Subsequently, anatomical tracing studies revealed strong connections between V1 and central V4 (Nakamura et al., 1993; Yukie and Iwai, 1985) and between V4 and TEO/TE (Desimone et al., 1980; Distler et al., 1993; Weller and Kaas, 1985). Further, V4 was found to have upper and lower field representations with identical functional properties using single electrode physiology (Gattass et al., 1988) and, recently, using retinotopic mapping using fMRI (Serenio et al., 1995) (see section 1.4.3 for more details on functional maps in V4.)

Form processing: Micro-electrode recording studies in V4 have found selectivity for oriented gratings (Desimone and Schein, 1987), contour curvature (Carlson et al., 2011; Gallant et al., 1993, 1996; Kobatake and Tanaka, 1994; Pasupathy and Connor, 1999, 2001), and surface texture (Arcizet et al., 2008; Hanazawa and Komatsu, 2001; Okazawa et al., 2015, 2017), suggesting that V4 plays a central role in the coding of object shape. Furthermore, V4 was demonstrated to be central to the conversion of a retinotopic coding scheme to an object-centered coding scheme (Pasupathy and Connor, 2001). As an example, V4 neurons respond consistently to curved gratings at many locations within their receptive fields (Gallant et al., 1996), while being sensitive to the relative position between a curved contour fragment and the rest of the object

(Pasupathy and Connor, 2001). In fact, this tuning for relative position remains intact in later areas (Brincat and Connor, 2004), suggesting that it is a fundamental transformation of visual representation. Recent reports have also attempted to distinguish between contour curvature defined by luminance, color, and textural contrast and have posited the theory that V4 neurons multiplex tuning for form, color and texture, perhaps reconciling earlier conflicting reports of color and form tuning and strengthening V4's functional role in processing object shape (Arcizet et al., 2008; Bushnell et al., 2011; Kim et al., 2019; Kosai et al., 2014; Okazawa et al., 2015; Oleskiw et al., 2018).

Disparity and depth processing: Binocular disparity is caused by the differences between the images formed on the two retinae and is the most sensitive cue for depth perception in primates (see figure 1.3A). At first, computation of disparity was thought to be the specialty of the dorsal visual pathway (Goodale and Milner, 1992; Livingstone and Hubel, 1988), but studies in V2, V4, and IT cortex have revealed that the ventral visual pathway plays a critical role in depth and disparity coding (von der Heydt et al., 2000; Hinkle and Connor, 2001; Janssen et al., 2000). The computation of the depth of an object relative to the viewer (absolute disparity) starts in early visual cortex but is enhanced in V4 (Hinkle and Connor, 2001; Watanabe et al., 2002). Additionally, using disparity cues, V4 neurons have been shown to have tuning for 3D oriented bars (Hinkle and Connor, 2005) (figure 1.3C), tilted and slanted planes, and curved planes (Hegd e and Van Essen, 2005; Tanabe et al., 2005).

While the conversion from absolute disparity (depth of object from viewer) to relative disparity (distance between objects in depth) starts in V2 (Cumming and Parker, 1999), a proportionally larger number of V4 cells show relative disparity tuning (Umeda et al., 2007), and micro-stimulation of V4 biases fine judgements of relative disparity (Shiozaki et al., 2012). Similar to conversion from retinotopic position to object-centered position, this transformation between

absolute to relative disparity serves to compact the diffuse representation of object features and makes them available for efficient read-out. Tuning for relative disparity enables fine judgements in shape disparity and serves 3D object representation (Shiozaki et al., 2012) (see figure 1.3D). This is distinct from tuning for coarse relative disparity found in the dorsal visual pathway (DeAngelis and Uka, 2003; Neri et al., 2004), which is important for visually-guided movements like reaching and grasping (Blohm et al., 2009).

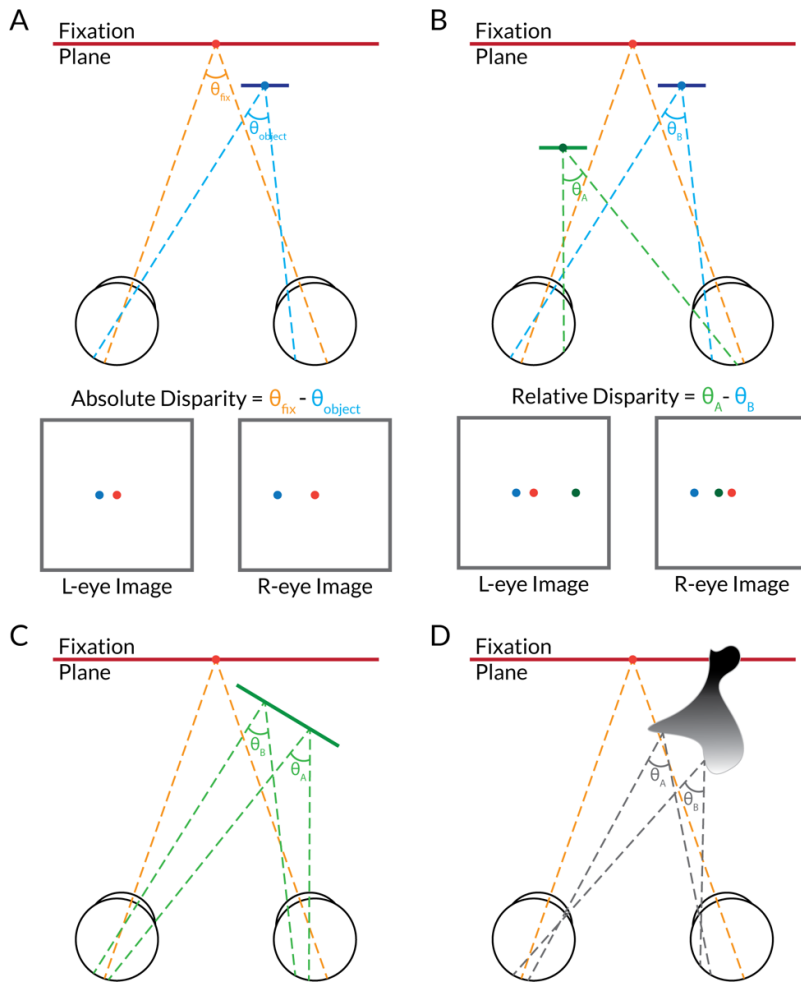


Figure 1.3: Absolute and relative disparity and their contribution to depth and solid form processing.

(A) The depth of a point in front of the fixation plane causes a disparity in the retinal images. This is called the absolute disparity of the point.

(B) The difference between the absolute disparity of two points is called the relative disparity between the points.

(C) The respective relative disparities of points along a line oriented in three dimensions can be used to calculate the 3D orientation of the line.

(D) The differences in relative disparities of the visible points on the surface of a solid shape in 3D space can be used to calculate the solid shape of the object.

1.3 Solid shape vision

The world observed to be ‘replete with objects’ in the first sentence of this introduction happens to be a 3D world with 3D objects. The solid shapes of these objects are informative about their physical affordances, surface smoothness, underlying material, and associated features like temperature, weight, balance, elasticity, etc. To interact with and make inferences based on these object properties, the solid shape of the object is extracted from the retinal image by transformations in the ventral visual pathway. Yet, most of the literature cited above has studied object vision in intermediate ventral visual stream with planar stimuli like sinusoidal gratings, filled or outline contour shapes, or angled bars and planes. This is consistent with the notion that the processing of solid shape from shading, motion, disparity, or texture either occurs in the dorsal visual stream (Tsutsui et al., 2005) or beyond V4, in IT cortex (see Orban, 2011 (Orban, 2011) for a review that dismisses V4 outright as a potential site for such transformations). Nevertheless, it is likely that the visual areas in both processing pathways interact directly or via parietal areas to achieve robust object vision (Janssen et al., 2018). The following paragraphs first offer a brief summary of the cues available to extract solid shape, before reviewing physiological investigations of solid shape representations in the ventral visual pathway starting at the level of V4.

1.3.1 Solid shape from various cues

Shading and specularity are the most prevalent cues for solid shape in the visual world. Humans robustly recognize solid shapes from these monocular cues (Todd and Mingolla, 1983).

Mathematically, determining shape from shading is known to be a problem with no unique solution. Natural image statistics can be utilized to narrow the space of solutions to a reasonable number (Barron and Malik, 2012) and addition of supplementary 3D cues (like disparity or motion) improves solid shape perception (Lee and Saunders, 2011; Norman et al., 2004).

Inferring solid shape from shading and specularities requires computation of the first (orientation) and second (curvature) derivative of surface contrast, respectively (Fleming et al., 2004). In fact, surface curvature-based representations of shape shading and contrast could be the mechanism underlying perceptual constancy of solid shapes (Todd, 2004).

Binocular disparity defined shape, also called stereopsis or cyclopean perception, requires knowledge of both the absolute (zeroth order) and relative (first order) disparity of the shape, as well as fine judgements of depth differences. Interestingly, much like how computation of shape from shading requires first- and second-order derivatives of contrast, calculating shape from disparity requires the computation of the second-order derivative of disparity, also called

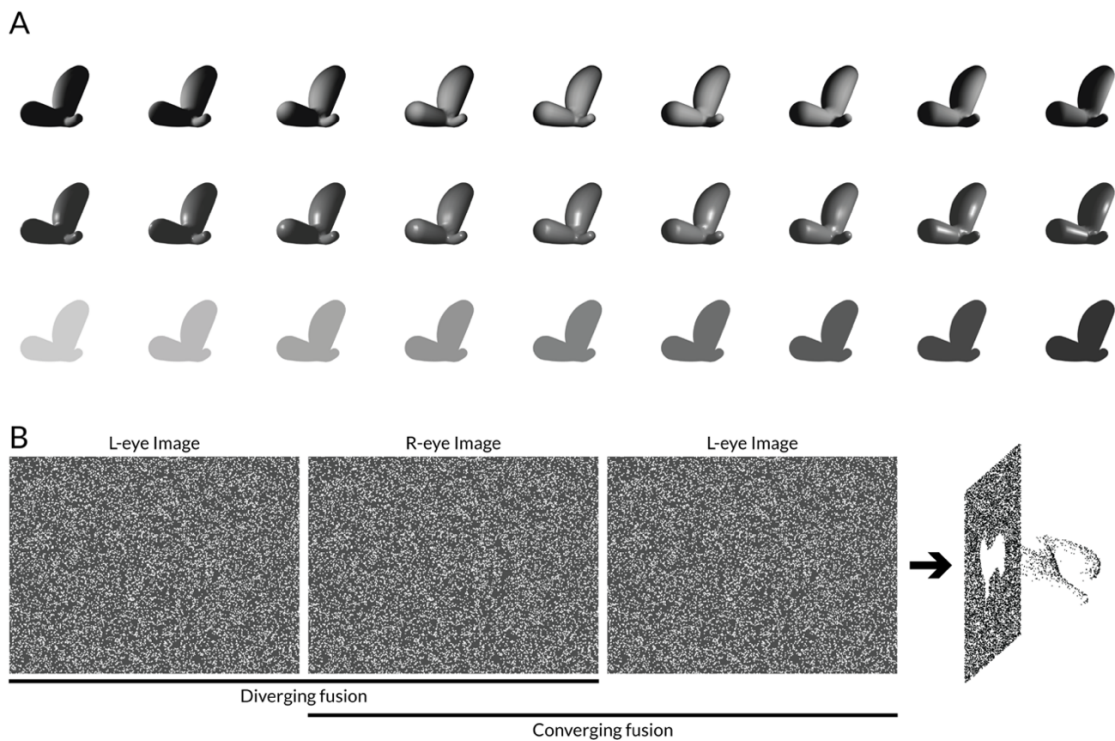


Figure 1.4: Shading and stereoscopic cues for solid shape perception.

(A) Matte (top row) and polished (middle row) versions of a solid shape presented in different lighting conditions (from left to right the light position changes in a circular path from the left to front to right). 2D planar silhouettes of the same shape are shown in decreasing luminance contrast (bottom row). (B) Stereoscopic images for a solid shape for diverging and converging fusion. A diagram of the perceived shape is shown on the right.

disparity curvature (Rogers and Cagenello, 1989). This is the key difference between depth perception and the perception of a solid shape from disparity – depth perception requires knowledge of the absolute disparity, whereas solid shape perception requires the processing of smooth changes in disparity (Kimmel, 2002).

Motion may also serve as a potent cue for solid shape inference. This is often demonstrated using so called random dot kinematograms (RDKs). In these stimuli, small dots are randomly positioned on a screen, and some of them are moved in a coherent manner about a shared axis. This results in the perception of a single (solid) cloud of dots rotating around the axis, instead of independently moving dots (Bradley et al., 1998; Siegel and Andersen, 1988). RDKs can be used to create vivid solid shape percepts and have been used to study shape-from-motion in humans (Kourtzi and Kanwisher, 2001; Murray et al., 2003; Treue et al., 1991) and monkeys (Burac^ˇas and Albright, 1996; Raiguel et al., 1995). Incidentally, the area in which motion signals for shape processing are integrated remains unknown, but both motion processing areas like MT/MST (in monkeys) and object processing areas like lateral occipital cortex (in humans) have been considered candidates in these studies.

1.3.2 Solid shape in V4

Given the representation of relative disparity, 3D orientation, contour curvature, and surface texture in V4, and the prerequisites of solid shape inference from various cues, V4 appears to be ideally situated for solid shape representation. This claim has some supportive evidence. Bilateral lesions in V4 caused monkeys to lose the ability to discriminate images of solid, shaded objects (Merigan and Pham, 1998). In human psychophysical experiments employing parallel visual search of solid/flat targets among flat/solid distractors, solid shapes pop out rapidly (< 80ms) (Enns and Rensink, 1990). This pop out perceptual effect is thought to be related to the computation of solid shape and reflectance in early visual cortex (Sun and Perona, 1996). Further,

V4 neurons are selective for convex or concave deformations of a 3D plane rendered in stereogram form (Hegd  and Van Essen, 2005). When spheres with concave and convex undulations (“potatoes”) rendered with shading cues were used, V4 neurons weakly, if at all, selected for the solid shape version over the planar silhouette (Arcizet et al., 2009), but the population of V4 neurons was overall significantly selective for the solid shape. Researchers have interpreted these somewhat conflicting results to mean that V4 is a critical stage in solid shape processing but doesn’t explicitly represent solid shape.

An alternative possibility is that V4 does indeed explicitly represent solid shapes, but that previous investigations were not able to sample a large enough shape space to identify solid shape processing. While the space of planar contour curvature can be exhaustively mapped by parametrically changing the contour curvature and orientation (Carlson et al., 2011; Pasupathy and Connor, 2001), this is not the case for solid shapes, which require many more parameters to be described completely. This dimensionality explosion makes it virtually impossible to parametrically search for tuning peaks for solid shapes. For this reason, the studies mentioned above only sampled parametric sub-spaces, selected by making *a priori* assumptions about the tuning space. However, the limited complexity used in these studies possibly prevented a true assessment of solid shape coding in V4. To truly measure preference for solid shape geometry, a search algorithm must be employed which systematically and adaptively samples the relevant space. This would result in a set of stimuli that elicit diverse responses from each neuron and can effectively constrain models of solid shape tuning. Additionally, it is imperative to validate solid shape preference and tuning against a large set of controls. Shaded shapes with and without specularities should be tested to check for first and second order shading preference. Multiple contrasts should be tested to rule out luminance tuning effects. Shapes must be rendered with monocular cues other than lighting, like texture, material, and optical properties. Solid and planar stereogram shapes must be tested at multiple absolute disparities to delineate disparity tuning and

shape tuning. Such a large battery of tests requires an effective and efficient experimental paradigm with affordances for long or short shape mapping experiments and prerequisite and post-hoc tests. Section 1.5 expands on the importance of studying solid shape coding in V4 and introduces the methods employed in this thesis.

1.3.3 Solid shape processing beyond V4

Higher-order ventral visual pathway areas in the inferotemporal cortex (specifically, TE) have been heavily credited for processing solid object shape. Indeed, the first results of solid shape processing in visual cortex were obtained in TE (Hung et al., 2012; Janssen et al., 1999, 2000; Vaziri et al., 2014; Yamane et al., 2008). TE has been shown to represent all surface orientations in the hemispherical domain nearest to the viewer, i.e. the orientations that have normal components in the direction of the viewer and therefore visible to the viewer (Yamane et al., 2008). TE neurons are also tuned to the surface curvatures of dimples, creases, bumps etc. of convex object shapes (Janssen et al., 2000, 2001; Yamane et al., 2008). Further, TE neurons were found to simultaneously represent the medial axis components of object shapes and the surface around the medial axis of the object (Hung et al., 2012). This representation was consistent across changes in lighting conditions, stereoscopic depth, large ranges of fronto-planar and out-of-plane 3D rotation, object scale, and across shading and stereoscopic cues. In addition to shape fragments, TE and TEO neurons are also known to represent the relative positions of specific shape fragments (Brincat and Connor, 2004; Hung et al., 2012). For instance, a TE neuron could be modeled as representing a linear combination of convexities and concavities at various locations on the surface of the shape. Such an explicit representation of solid objects operates in the 3D space of physical structures and structural relationships and not in retinotopic coordinates. On even larger scales, geometric features belong not to solid shapes but to visual scenes like large, smooth gradations in the ground plane, tilts and angles of walls, and crevasses and hills.

Indeed, TE neurons have been found to represent such complex scene features (Vaziri et al., 2014).

1.4 Micro-organization of visual areas

The study of the micro-organization of visual cortical areas is inexorably linked to the understanding of coding transformations. Often, cortical maps of stimulus characteristics have been found for the tuning dimensions that are emergent in an area, and those considered central to processing in that area (see examples below). Knowing the cortical map of a region can thus provide crucial insights into its critical dimensions. Why and how the layout of stimulus characteristics aids coding transformations within the area, or in areas up and down the visual processing hierarchy, and what, ultimately, is the need for a *cortical map* at all, are questions of critical importance. The following sections attempt to briefly outline known cortical maps, the current understanding and utility of cortical maps, and why it is vital to investigate the mapping of solid shape characteristics in visual cortex.

A vast majority of studies of functional organization have targeted V1. As a result, most theories of micro-organization are based on cortical maps in V1, thalamo-cortical inputs to V1, intra-columnar connections, and feedforward connections to higher visual areas. To better appreciate models of micro-organization and theories of cortical clustering, a small survey of cortical maps in V1 follows in the first section. The next section evaluates the utility of cortical maps. The final section lists the known functional maps in higher visual areas and posits a map for solid shape preference and solid shape fragments.

1.4.1 Functional organization and maps in primary visual cortex

Due to the stereotypic and orderly axonal projections from LGN to V1, a *retinotopic map* is produced (Connolly and Van Essen, 1984; Tootell et al., 1982; Udin and Fawcett, 1988; Van Essen and Maunsell, 1980) which has most recently been visualized in fMRI scanners (Lafersousa and Conway, 2013; Sereno et al., 1995). The representations of the visual vertical and horizontal meridian have served as boundaries between visual areas for decades (Engel et al., 1997). Retinotopy can be thought of as the biological epiphenomenon of organized retinal and LGN projections.

The more striking observation, first made in somatosensory cortex by Mountcastle, is that the *functional* properties of neurons are organized in cortical columns “intermingled in a mosaic-like fashion” (Mountcastle, 1957). Mountcastle observed that the tuning functions of vertically aligned neurons in cat somatosensory cortex remained largely similar. Since known as “columnar organization”, it was demonstrated by Hubel and Wiesel to hold true in cat primary visual cortex for the representation of orientation and ocular dominance (OD) (Hubel and Wiesel, 1962). If an electrode is inserted horizontally instead of vertically, peaks of orientation tuning functions transition smoothly between orientations. This organization represents a “functional cortical map” i.e. a map of neural tuning properties across the cortex parallel to the pial surface. Columnar organization and cortical maps together describe the 3D layout of neurons in cortex. It is worth noting that while the terms are often used interchangeably, neither form of organization is actually necessary for the other to exist. The next few paragraphs list the known functional maps in V1 and, where relevant, their mutual alignment relative to the map of orientation preference.

After injecting radioactive tritiated hydrogen into one eye, Hubel and Wiesel observed a map of OD in cat V1, arranged as interdigitated stripes of left- and right-eye selective regions. In additional experiments using microelectrode recordings in monkey V1, they also showed that

orientation preference remains constant as the electrode is advanced perpendicular to cortex, and changes systematically as the electrode is advanced tangential to the surface of cortex (Hubel and Wiesel, 1968; LeVay et al., 1975). More recently, optical imaging of intrinsic signals (difference between the reflectance between oxygenated and deoxygenated blood) and of signals from voltage-sensitive dyes has been used to visualize 2D orientation pinwheel maps and recapitulate OD stripes (Blasdel and Salama, 1986; Grinvald et al., 1986; Ts'o et al., 1990). Further, these and other experiments have demonstrated the orthogonal relationship between orientation and ocular dominance maps (Bartfeld and Grinvald, 1992; Crair et al., 1997) in which boundaries between iso-orientation columns tend to intersect ocular dominance boundaries at 90° . Other low-level stimulus parameters like color (Landisman and Ts'o, 2002; Ts'o and Gilbert, 1988), spatial frequency (Issa et al., 2000), and spatial phase (Aronov et al., 2003) have also been mapped with optical imaging and microelectrode recording techniques. Maps of second-order stimulus features like disparity (Kara and Boyd, 2009) and direction (Shmuel and Grinvald, 1996; Weliky et al., 1996) have been observed in cats and ferrets but not monkeys.

Cortical maps of stimulus properties like orientation and ocular dominance should not be studied only in isolation – their mutual relationships produce key insights into cortical computations (Kremkow et al., 2016; Sirovich and Uglesich, 2004). Recently, maps of spatial frequency (SF) have been discovered to be orthogonal to orientation maps with two-photon calcium imaging (Nauhaus et al., 2012). While SF was known to be organized in a columnar fashion and this organization was observed to be patchy (Born and Tootell, 1991; Issa et al., 2000; Silverman et al., 1989; Tootell et al., 1988), single cell resolution was required to investigate the efficient joint representation of these parameters. An example of non-orthogonal organization was revealed when researchers mapped orientation, OD, and SF together and found that boundaries of iso-orientation domains intersect both OD and SF domains at 90° , and therefore the SF and OD domains run parallel to each other (Nauhaus et al., 2016); monocular zones are centered on low

SF peaks and binocular zones at high SF peaks. Importantly, it was additionally observed that the receptive field positions of neurons in V1 regions as large as $200\mu\text{m} \times 200\mu\text{m}$ had negligible scatter, which places strong constraints on how tightly different features maps have to be aligned with respect to each other and the retinotopic map to achieve complete coverage of all stimulus features at every location in visual space. Most recently, the joint representation of orientation and color in V1 has been investigated (Garg et al., 2019). Contrary to earlier studies proposing a strict segregation of orientation and color processing (Livingstone and Hubel, 1988), this most recent study observed a spectrum of processing of both features, with some neurons encoding either feature independently, and other neurons processing them simultaneously. The joint organization of orientation and color showed some alignment to the cytoarchitecture of V1, which is also thought to be linked to other maps like SF. Thus, while it remains to be studied, it seems likely that V1 contains an overarching orderly architecture for all encoded features.

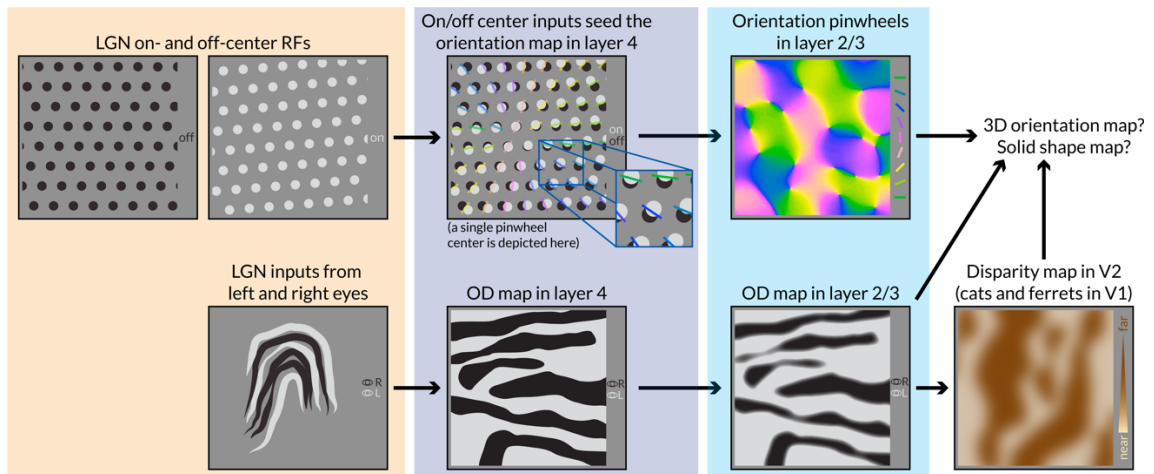


Figure 1.5: Functional maps in V1 and pooling hypotheses.

LGN: Inputs from left and right eyes are segregated. Neurons have either on- or off-center RFs. Slight distortion or bias in the alignment of on- and off-center neurons creates a dipole to form in the input layers of V1.

V1 layer 4: The moiré pattern formed by offset on- and off-center inputs seeds the orientation tuning map in V1. Also, a segregated targeting of contra- and ipsilateral eye inputs causes ocular dominance bands.

V1 layer 2/3: Pooling layer 4 inputs causes the formation of binocular zones between monocular zones.

Other maps: It is conceivable that maps for disparity, color, direction, spatial frequency, spatial phase etc. are formed because of similar pooling mechanisms but more experiments are required to corroborate this hypothesis.

1.4.2 The utility of cortical maps and their mutual alignment

Researchers have put forth many hypotheses as to why cortical maps in V1 are laid out in a stereotypic fashion (Horton and Adams, 2005; Nauhaus and Nielsen, 2014; Van Hooser, 2007) but the fundamental reason behind the existence of maps is not obvious. Gray squirrels, for example, lack systematic orientation maps in primary visual cortex and yet possess neurons that are just as effectively tuned to orientation as those in primate V1 (Van Hooser et al., 2005). So, why should cortical maps exist at all? The following hypotheses attempt to address this conundrum. At least in primates, maps may allow fast, local computations of visual features that can serve to conserve energy. Additionally, an orderly organization of tuning properties across neighboring neurons would allow pooling mechanisms that operate indiscriminately on all neurons within a local region to extract meaningful new tuning. The same pooling mechanism applied to randomly organized neurons would suffer from conflicting responses of the participating neurons. More generally, addressing neurons with related tuning properties can be achieved using a simple spatial mechanism for orderly feature maps, but requires very specific rules for random tuning organization (Nauhaus and Nielsen, 2014).

Striking evidence for the formation of V1 orientation-tuned columns (not the orientation map) indeed hints at the role of spatial pooling mechanisms in the emergence of new tuning properties (Jin et al., 2011). Researchers recorded single LGN neurons and then identified the V1 column that those neurons projected to. This revealed that the receptive fields of on- and off-center thalamic inputs to layer 4 of a single V1 column overlap. Crucially, while little structure was apparent in the converging on- and off-center inputs when considered individually, a simple spatial pooling algorithm resulted in an input population receptive field with an orientation that closely resembled that of the V1 column (as depicted in the magnified inset in figure 1.5B). A similar pooling concept has been applied to feedforward connections from V1 to V4. By simply

pooling a sample of inputs from a circular region of V1 researchers could model the receptive field size and shape of V4 neurons without need for any extraneous magnification factor (Motter, 2009).

There is also strong interest in how the overall layout of the orientation tuning map could emerge based on the spatial structure of the retinal input in V1. One mechanistic theory behind the emergence of orientation maps is based on the observation that retinal ganglion cell (RGC) mosaics have a stereotypical relative alignment (Paik and Ringach, 2011). Because of this alignment, neighboring on- and off-center RGC create ‘dipoles’ with orientations that vary in a systematic spatial pattern, which, in turn, can be related to the layout of orientation maps in V1 (see top row in figure 1.5). Given a stereotypic arrangement of RGCs, this model predicts the formation of a hexagonal pattern of iso-domain orientation columns on the cortical surface, which is indeed observed in cat, ferret, monkey, and tree shrew V1. Similar pooling principles could potentially be applied to create models of OD, SF, and disparity maps, and these models could also be extrapolated to explain the emergence of solid shape processing, for instance, in higher visual areas.

As mentioned above, the mutual arrangement of functional maps on the cortical surface is surprisingly stereotypic. As elaborated in the next paragraph, the leading and best corroborated theory of *why* functional maps are co-organized in such a tightly stereotypic fashion is based on ‘dimensionality reduction’ of the neural response space to uniformly cover stimulus space (Durbin and Mitchison, 1990). Briefly, two points in stimulus parameter space should ideally be placed close-by on the cortical sheet and therefore use *minimal wiring*. Another way to approach this problem is in terms of the ‘travelling salesman problem’ where an algorithm tries to *minimize the path* of a salesman (traversal on the cortical sheet) as he travels between cities (stimulus parameters). The latter approach amounts to reducing the dimensionality of response space and

mapping it to stimulus space (Durbin and Mitchison, 1990; Goodhill and Sejnowski, 1997).

Before elaborating further, this theory works under the following assumptions: (a) all visual areas have a single, continuous retinotopic map on which smooth, periodic maps of stimulus characteristics are overlaid, (b) these maps are often structurally related, (c) all reasonable permutations of stimulus parameters should be represented, and (d) most cortical operations are performed locally amongst neurons with similar tuning characteristics.

Following from (d) above, the critical optimization would be to minimize wiring/axonal length between like neurons, which restricts neurons with similar tuning characteristics to be closer together. In effect, the solution to the wiring minimization problem is to place neurons with similar tuning next to each other (Koulakov and Chklovskii, 2001). This solution works if the dimensionality of stimulus space matches the dimensionality of the map. Since the cortical surface is 2-dimensional and the stimulus parameter space is often larger than 2-dimensional (orientation, SF, phase, color, disparity, etc.), a simple wiring minimization algorithm will not be able to model cortical maps. Alternatively, if the dimensionality of neural responses is reduced to two dimensions and mapped onto stimulus space, effectively reversing how the wiring problem is solved, all tuning dimensions can be mapped onto the 2D cortical sheet (Durbin and Mitchison, 1990). This dimensionality reduction and mapping procedure has been shown to produce orientation and OD maps that resemble observed maps in V1. In fact, this study also revealed that such a model has many potential smooth mappings, which include the solutions with minimum wiring. In addition to orientation pinwheels, the model captures ‘orientation fractures’ – regions of discontinuities in the orientation map (Blasdel and Salama, 1986). One interesting prediction of this model was that, because retinotopy was treated similarly to orientation (both free parameters in the model), the receptive field location will change smoothly and rapidly along the orientation fractures and slowly everywhere else. The large fractures have been demonstrated in physiological studies in cat (Das and Gilbert, 1997) and the tight relative alignment of RFs within

orientation domains have been shown in recent studies of functional retinotopy using electrophysiology (Jin et al., 2011), optical imaging (Blasdel and Campbell, 2001) and two-photon imaging (Nauhaus et al., 2016).

Lastly, Swindale and colleagues have proposed a quantitative measure of the uniformity of stimulus parameter coverage (Swindale et al., 2000). Using this metric, they demonstrated that if the mutual configuration of the functional maps is perturbed during calculations, the ability of cortex to cover all permutations of stimulus parameters reduces. Thus, they concluded that, at least in the cat, the functional maps are mutually structured in a very precise way in order to improve stimulus coverage. If there are many more stimulus parameters mapped in a single area, coverage would naturally reduce, but efficient tiling of stimulus space might be a way to achieve adequate coverage.

1.4.3 Functional maps in higher ventral stream areas

The theories listed above largely centered on explaining cortical maps and columnar organization in V1. While the literature on cortical maps in higher visual areas is sparse and more controversial, the next few paragraphs briefly survey the known maps and, where possible, hypothesize how they might form based on pooling mechanisms described above. The fundamental hurdle in discovering maps in higher visual areas is that the critical coding dimensions in these areas are largely unknown. As discussed in the previous section, V2 and V4 neurons exhibit complex responses to color, disparity, orientation, etc. Also, the further downstream an area, the larger the dimensionality of coding transformations, and less clear how hypotheses of cortical maps should be constrained. Researchers therefore form hypotheses based on anatomical connectivity patterns between areas and the complexity of responses of single neurons to stimuli.

In V2, the known maps are often aligned with regions revealed by staining for cytochrome oxidase, a mitochondrial enzyme which is a marker of enhanced cortical activity (but see (Takahata, 2016)). This staining reveals ‘blobs’ in V1 and thin, thick, and pale ‘stripes’ in V2 (Livingstone and Hubel, 1984; Tootell et al., 1983). Physiological studies in these anatomically defined regions found unoriented, color-selective, monocular cells in V1 blobs and V2 thin stripes, and oriented, disparity-selective cells in V1 interblobs and V2 thick stripes. In V2, optical imaging and targeted microelectrode recordings have uncovered cortical functional maps aligned to thin CO stripes for color/hue (Lim et al., 2009; Shipp and Zeki, 2002; Ts’o et al., 2001; Xiao et al., 2003) and maps aligned to thick CO stripes for orientation (Malach et al., 1994; Ts’o et al., 2001), horizontal disparity (Chen et al., 2008), motion-defined orientation (Chen et al., 2016) and motion direction (Lu et al., 2010).

In V4, CO does not stain regions of high or low cortical activation. However, retrograde tracers, when injected into V4, labeled neurons from thin but not thick stripes in V2 and, when injected in posterior inferotemporal cortex (PIT or TEO), labeled patches of neurons in V4 (Nakamura et al., 1993). Also, whereas central V4 receives direct projections from V1 and is connected to inferotemporal cortex (Shipp and Zeki, 1995; Zeki, 1969), peripheral V4 is connected more definitively to parietal cortex areas like LIP and MST (Baizer et al., 1991). Tacitly, these connectivity patterns suggest differences in the functional properties of neurons in different parts of V4, and possibly the functional organization. Most studies therefore target foveal and parafoveal V4 to find representations pertaining to object processing.

Optical imaging-targeted electrophysiology experiments in central V4 found “modules” for form processing in which neurons were tuned for the same orientation (Ghose and Ts’o, 1997). There were also larger modules that contained neurons suppressed by large stimuli, which, in turn, contained orientation domains. Similar experiments with colored gratings revealed organization

for color, brightness, and contour orientation (Kotake et al., 2009; Tanigawa et al., 2010). These results show a remarkable juxtaposition between domains of color selectivity and orientation preference.

This color-form segregated model is also corroborated by fMRI experiments that divide V4 into highly color-selective, unoriented regions (“globs”) arranged chromatopically with orientation selective domains juxtaposed alongside them (Conway and Tsao, 2006; Conway et al., 2007).

The strictly segregated processing of color and orientation is a radical model of coding transformations in V4, given the mixed selectivity of similar domains in V2 and optical imaging experiments outlined above. To what extent these regions correspond to complex shape properties is yet unknown. Nevertheless, these studies contribute evidence to support patchy organization of shape features in V4.

1.5 Studying solid shape representation in V4

Fundamentally, the visual information in the physical world and our perception of it is not divided into texture, disparity, color, motion, etc. It is divided into objects with three-dimensional form with which we learn to interact. The evidence for solid shape processing in V4 has been scant and controversial (Orban, 2011; Orban et al., 2006). Still, theories of the origin of solid shape representation in TE strongly implicate V4 as a potential candidate. This is further supported by the observation that the planar orientations and curvatures represented in V4 reflect the derivatives of 3D surface contrast in real world objects and scenes. Therefore, an investigation of whether or not V4 computes properties of solid shape fragments and in turn supports representation of higher-level geometries in later ventral visual pathway areas would be central to the understanding of object perception and recognition. The results in this thesis directly address these open problems.

In experiments described in chapter 2, an adaptive shape morphing algorithm was used to sample responses to solid shape in V4 and compare them against corresponding planar shapes and other controls. This method efficiently sampled a potentially infinite stimulus space by evolutionarily morphing solid shape stimuli based on neural responses. Later, response-weighted averaged matrices were used to precisely describe the tuning peak for shape fragments in three-dimensional space. These data suggest that solid shape fragments are the elementary building blocks of object vision, not spatial frequency patterns or other image properties.

While there have been efforts to find cortical maps for two-dimensional contours (Ghose and Ts'o, 1997), and disparity-defined edges (Fang et al., 2018), and color/hue (Kotake et al., 2009; Li et al., 2014), two major hurdles in the path of the discovery of an organizational motif in V4 are the development of fast, high resolution imaging techniques, and computational techniques to effectively sample the space of solid shape stimuli. Advancements in two-photon calcium imaging have made the technology ubiquitous in visual system research, but imaging in higher-order areas in monkeys poses a unique set of technical challenges – imaging for long periods of time without compromising image quality and sensor expression and, in the case of anesthetized experiments, adjusting the depth of anesthesia precisely to prevent the dampening of visual response to complex stimuli. More importantly, finding tuning for solid and flat shape features requires an ability to probe the infinite space of possible solid and flat shapes effectively. These experiments have to be performed under strict hypotheses about coding transformations; they provide a framework without which it would be extremely challenging to interpret neural responses and cortical maps.

In experiments described in chapter 3, a parameterized solid shape test was devised for two-photon imaging experiments to study the micro-organization in V4. This approach was chosen

over an adaptive algorithm (as in chapter 2) to optimize imaging time and maximize explanatory power across imaging regions. These data revealed that V4 neurons are clustered based on their preference for solid versus flat shape fragments. This points again to solid shape being the elementary coding motif in V4 as opposed to image-level properties.

2 Solid shape representation in V4

This chapter contains a subset of the results currently submitted as a journal manuscript. If accepted, it will appear as “Srinath, Emonds, Wang, Nielsen, Connor. (2019). Volumes, not patterns, are the basic elements of object vision”.

2.1 Introduction

Information about the solid geometry of objects is lost in two-dimensional retinal images and must be recovered across multiple processing stages in the ventral visual stream (Van Essen et al., 1992). The hierarchy of coding transformations implemented in the ventral visual stream helps us make inferences about object identity and form and facilitates discrimination of and interaction with visual objects. Models of object processing in early and mid-level cortical stages have focused on transformations in the 2D domain. For instance, neurons in a mid-level processing stage, area V4, have been shown to encode image properties, including contour orientation and curvature, spatial frequency, and texture. Representation of these features supports the hypothesis that ventral visual stream areas successively compute various image transformations to reconstruct object boundaries, surface textural patterns, and relationship to other objects. The results described in this chapter contrast this theory by outlining evidence that shows that neurons in V4 explicitly encode solid shape geometry rather than planar silhouettes and texture patterns. Additionally, this representation was found to be consistent across cues for 3D reality like shading, specularity, binocular disparity, reflectivity, and refractivity. This encoding is, therefore, highly abstracted from the encoding of planar images, underlies the physical structure of solid shapes, and is not describable in terms of image transformations. These results argue that, even in the earliest stages of the ventral visual pathway, coding transformations are optimized for the extraction of 3D geometric reality rather than the computation of image features.

2.1.1 Experimental approaches to studying shape coding

Most of the investigations of coding transformations in the ventral visual pathway can be classified based on stimulus set type and explanation type. Broadly, studies utilize reductive, exhaustive, parameter-agnostic, or response-adaptive stimulus sets. Also, based on the experiment, models of V4 responses could contain units that are constrained to be tuned to a set of manually specified parameters. These categories are a result of technical constraints and differing philosophies of the utility and purpose of coding transformations.

Studies utilizing a reductive stimulus approach start with a large stimulus set – a set of complex shape outlines, for instance – and based on the neural responses, attempt to iteratively mask the irrelevant portions of the stimulus and discover the simplest feature that drives the neuron. This approach assumes that the ventral visual pathway is a series of feature detectors, where each neuron is tasked with detecting a single feature in the stimulus. Accordingly, this technique was first used to identify that the ventral visual pathway contains neurons with preferences for complex shape features (Kobatake and Tanaka, 1994). More recently, a modified version of this approach was used by Jim DiCarlo and colleagues (Bashivan et al., 2019) where responses to a large set of natural images was used to generate a simple texture using a deep neural network model of V4 which produced a higher firing rate than any of the tested images. Tacitly, this approach produces a qualitative description of neural responses, biased by the chosen stimulus set and the feature reduction procedure.

The second category of experiments assumes the coding domain of the area in question – orientation and spatial frequency for V1, curvature for V4, faces for face patches, etc. – and exhaustively samples the relevant stimulus space. Unlike the previous approach, which produces the estimate of a ‘critical feature’, these experiments produce a quantitative estimate of the neural tuning manifold. In V4, this approach has been successful by assuming encoding of spatial

derivative-based features (Gallant et al., 1996), curvature-based features (Pasupathy and Connor, 1999, 2001; Yau et al., 2013), 3D orientation (Hinkle and Connor, 2001, 2002), color (Zeki, 1983), etc. In order to be tractable, these experiments either assume a simple coding domain or very coarsely sample a complex domain, ignoring factorization of multiple parameters.

Additionally, a neuron that seems to be tuned to the variance in a certain stimulus parameter could actually be tuned to a correlated parameter not being varied in the experiment. In areas where researchers had a well-constrained hypothesis of the coding domain, this approach has been very successful in finding the critical features and tuning. Although some models with unconstrained parameters have effectively fit neural responses from these experiments (Cadieu et al., 2007; Popovkina et al., 2019), they are more amenable to models which contain units that are tuned to the dimensions which were systematically varied during the experiment.

The third category of experiments uses parameter-agnostic stimulus sets – either natural images or random pixel patterns. These stimuli are unconstrained by geometric parameters and often have the advantage of being more similar to natural visual experience. This approach is generally paired with a complex model with a large set of parameters, which is used to fit responses to these natural images. Examples of this include deep network models with millions of free parameters and unconstrained units (Yamins et al., 2014) or HMAX with hundreds of free parameters and tuning constrained units (Riesenhuber and Poggio, 1999). Gallant and colleagues pioneered this technique and others have used it recently to predict V4 neural responses successfully (Abbasi-Asl et al., 2018; Cadieu et al., 2007; David et al., 2006; Pospisil et al., 2018; Yamins et al., 2014). While these results are impressive, because the deep network models don't contain units that are constrained to be tuned to a set of specified stimulus dimensions, it is generally very challenging for these studies to infer neural tuning, and, therefore, they fail to offer satisfactory coding hypotheses.

The response-adaptive stimulus set approach was pioneered by Connor and colleagues to study solid shape coding in anterior IT cortex (Yamane et al., 2008) and later to study the sparse coding of curvature in V4 (Carlson et al., 2011) and other aspects of shape tuning in the ventral visual pathway (Hung et al., 2012; Vaziri et al., 2014). This approach uses a stimulus generation algorithm that evolves stimuli based on the neural response as the experiment is being conducted. As a result of this effective sampling, using this approach can hugely increase the parameter space, while still retaining the ability to explain neural tuning in terms of geometrical parameters with tuning models. The critical drawback of this technique is that the stimulus generation and morphing algorithms make assumptions about the coding domain much like the parametric sampling technique, but because of the efficient sampling, the sampled space can be huge, even covering domains that are suspected to be coded in downstream areas. Also, because this technique employs searching for the tuning function, it is prone to find local minima in the search space and be affected by the initial conditions. Therefore, multiple independent searches must be performed to validate the tuning functions and, even when the search space is narrowing, the space must continuously be randomly sampled to provide novel starting conditions.

2.2 Results

2.2.1 2D and 3D shape information in area V4

Electrophysiological recordings were performed with single electrodes in V4 neurons in two rhesus macaque monkeys. During each recording session, the electrode was lowered into dorsal V4, and a single neuron was isolated. The receptive field location and size and the neuron's color preference were determined by drifting 2D filled contour shapes under the experimenter's control. The adaptive algorithm began with a set of 80 randomly constructed solid shapes. The random shape generation algorithm used here was adapted from (Hung et al., 2012), simplified for V4 response properties (see methods section 2.2.4 for details). These stimuli were randomly

divided into two lineages for later comparison and cross-validation. Each stimulus was presented for 750ms with a 250ms inter-stimulus interval. The response of the neuron to each stimulus was calculated as the spike rate during the stimulus presentation period. “Gen 1” in figure 2.1A depicts the first generation of stimuli of an example experiment arranged in decreasing order of neural response. The intensity of the red border around each object indicates the response of the neuron to that object.

Each subsequent generation (Gen 2-7) was created by partially morphing ancestor stimuli selected from previously run generations. The neural response to stimuli served as a fitness metric for the selection of ancestors. Stimuli were sampled from low, medium, and high response ranges to evolve diverse stimulus trees, each of which contain stimuli that elicit a wide range of responses. The probability of a stimulus producing descendants was proportional to its neural response.

In addition to morphed descendants, five high response ancestors were also selected from each lineage to be rendered in two solid and two planar conditions – solid matte- and polished-surface stimuli and planar high- and low-contrast stimuli. Planar stimuli were silhouettes of the ancestor shape and thus corresponded to the self-occlusion boundary of a naturalistic solid object instead of a randomly generated planar object. This gave us the ability to directly compare the responses between solid shapes their planar silhouettes, and also investigate neural tuning. This was done in parallel to the adaptive algorithm to continuously provide evidence for solid shape preference, instead of after the experiment as a post-hoc test. In figure 2.1, “Gen 2” exemplifies a full generation with twenty morphs (figure 2.1B) and twenty ‘preference-testing’ tetrads (figure 2.1C). “Gen 3-7” depict top morphed and tetrad responses (figure 2.1D). In this case, the responses to solid shapes were consistently higher than those to planar shapes.

To quantify this preference, an index for solid shape preference (SP) defined as:

$$SP = \frac{R_{solid} - R_{planar}}{\max(R_{solid}, R_{planar})}$$

where R_{solid} is the average response to matte- or polished-surface solid shapes, whichever is higher, and R_{planar} is the average response to low- or high-contrast planar shapes, whichever is higher. This score varies from -1 (purely planar-preferring neuron) to 0 (no preference) to 1 (purely solid-preferring neuron). For this neuron, $SP = 0.54$.

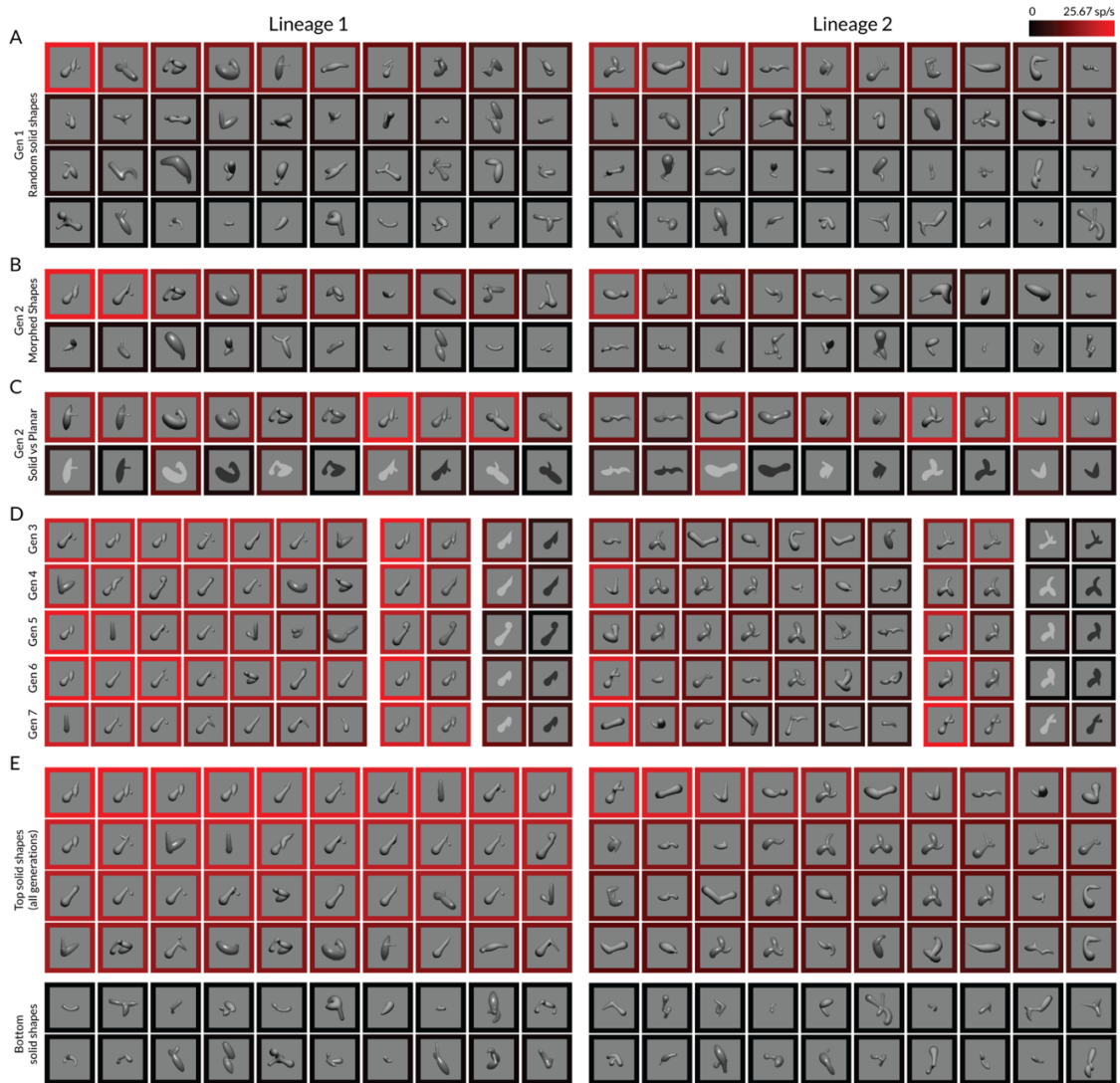


Figure 2.1: Adaptive solid shape sampling experiment.

Responses of a single V4 neuron to solid shape stimuli indicated by the color intensity (see scale bar, top right) of the image border for shapes evolving in two independent lineages (left and right columns). (A) Initial generation of 40 random stimuli per lineage sorted in descending order of trial-averaged responses. (B) Second generation of 16 partially morphed stimuli, selected probabilistically from generation 1, and 4 novel random shape stimuli. (C) 5 top response stimuli from the previous generation rendered as matte and specular solid shapes and high and low contrast planar (silhouette) shapes presented as tetrads. (D) Top adaptive algorithm stimuli and tetrads for generations 3-7. (E) Top 40 and bottom 20 stimuli across all generations depicting convergence across lineages and dense sampling of stimulus space in the high response range.

2.2.2 Parameterizing shape space and visualizing shape tuning

It is visually apparent that similar solid shape properties re-occur across the top stimuli in both lineages (figure 2.1E “top” responses). In this example, solid shapes that contained a straight shaft angled at 45° towards the lower left of the object elicited the highest responses. These shafts typically terminated in a bulbous end. To quantify and visualize neural tuning for these geometric properties, each shape must be parameterized in terms of the geometry of its parts. Then, based on the response of the neuron to each shape, shape parts can be weighted, and the space can be visualized.

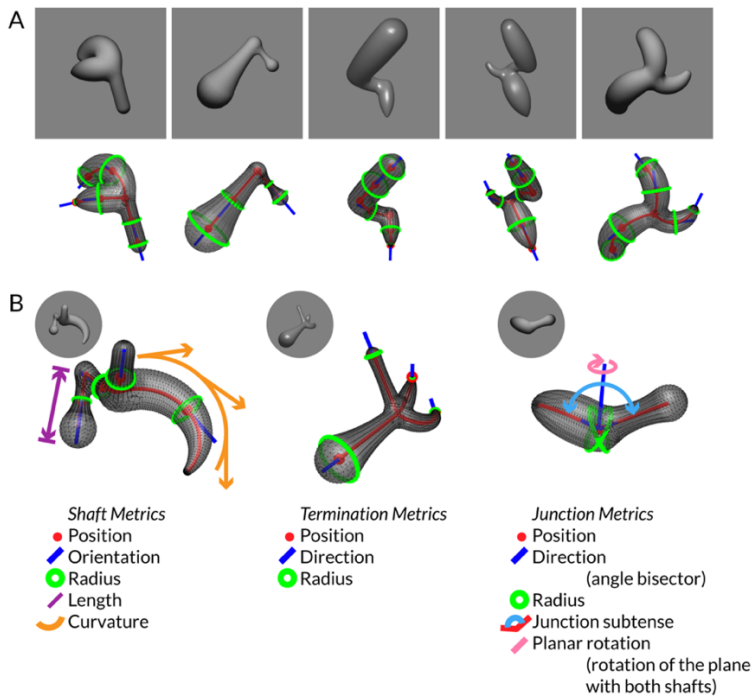


Figure 2.2: Metric shape analysis.

(A) Solid shape examples (top) and their respective shape parameterizations (bottom). Red axis indicates the medial axis of the shape, green circles indicate the width of each limb at critical locations along the limb, and blue lines indicate the tangents of the limbs at the same critical locations. (B) Various geometric parameters used to characterize the shapes. Shafts, junctions, and terminations are treated independently and therefore have independent RWA matrices. Position is 3-dimensional vector and orientation is a 3D unit vector, both represented in spherical coordinates. Planar rotation is a circular variable.

To this end, each object was divided into three shape fragment types – terminations, shafts, and junctions (figure 2.2). The contribution of these shape fragments to solid shape tuning was analyzed individually. Each of these parts was parameterized with dimensions for object-relative position, orientation or direction, length, radius, curvature of medial axis shafts, and their surrounding surfaces. In addition, junctions were parameterized by their subtense angle, the direction of the angle bisector, and the planar rotation about the angle bisector. Therefore, each shape was defined as a collection of points in three distinct geometric subspaces viz. shafts, junctions, and terminations. These points were then convolved with a Gaussian function weighted by the response of the stimulus (figure 2.3A). The response-weighted surface for each stimulus

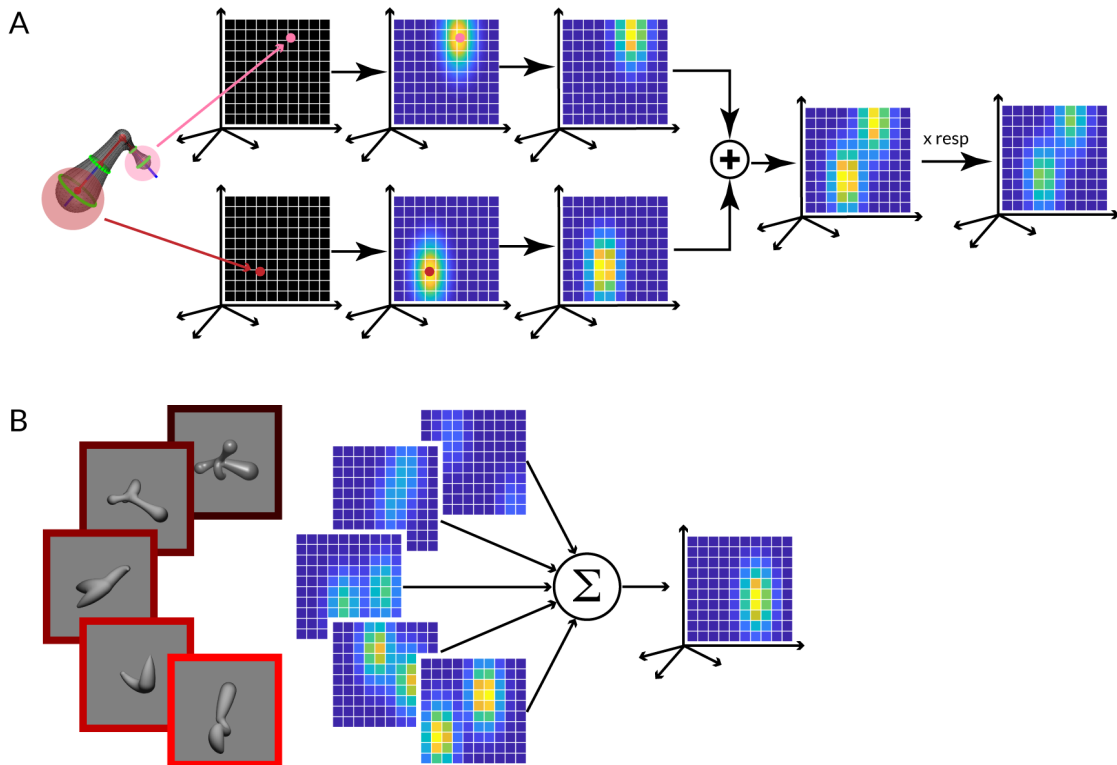


Figure 2.3: Response-weighted averaging (RWA) analysis.

(A) A stimulus with two terminations will have two ‘points of interest’ in the 6-D termination RWA matrix. Two gaussians are summed into an empty matrix and multiplied by the neural response to that stimulus. The value in each bin is dependent upon the Euclidean distance (in 6-D) between the bin center and the ‘point of interest’. (B) Each stimulus contributes one 6-D matrix which are all summed to produce the RWA matrix. This matrix is divided element-by-element by an unscaled matrix (not normalized by the neural response) to control for sampling biases (not shown in the schematic).

was summed and normalized by an unweighted version of the same matrix to correct for uneven sampling due to correlations in solid shape statistics (figure 2.3B). The resultant matrix was the response-weighted average (RWA) matrix. These calculations were performed separately for each independent lineage and then, to remove spurious peaks, multiplied to obtain the final RWA. Because the lineage RWAs were multiplied and not summed, a poor prediction would be an indication of either poor parameterization or lack of lineage convergence. To ensure that this multiplication still retained the ability to explain the neural tuning, the neural responses for the full set of stimuli (both lineages) were predicted using the final RWA.

The final RWA for the neuron in figure 2.1 is visualized in figure 2.4 as slices through the peak of the RWA, parallel to one or two geometric dimensions. This visualization clearly shows tuning to termination and shaft dimensions. As described above, the positional tuning for terminations and shafts was centered towards the bottom left of the object. Tuning for the direction of the termination and the orientation of the shaft coincides along the left-down axis. Shafts, unlike terminations, don't have an inherent direction and are depicted in a hemispherical space facing the viewer. Tuning for a large termination radius indicated a preference for bulbous terminations, whereas tuning for shaft radius was broad with a tendency towards slimmer shafts. This analysis adequately captures the variance in neuronal responses with respect to geometric shape dimensions.

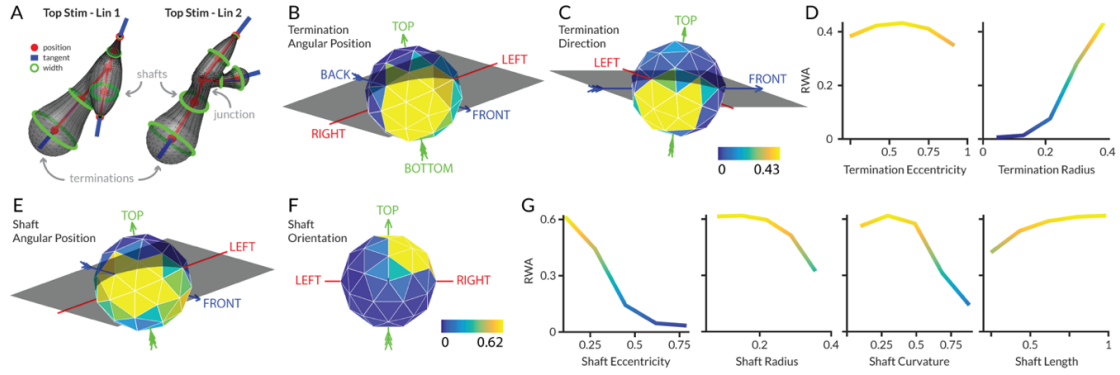


Figure 2.4: RWA matrix for an example neuron.

(A) Metric shape parameterization of top stimuli from both lineages (B-D) Slices of the lineage-product RWA through the neural tuning peak parallel to 1 (line plots) or 2 (icosahedrons) dimensions indicating tuning for bulbous terminations at the bottom-left of the shape in the bottom-left direction. (E-G) Same as (B-D) for 3D oriented shafts showing tuning for straight shafts situated to the left-back-bottom of the center-of-mass of the shape oriented along the bottom-left (or top-right) axis.

More examples of solid shape tuning are presented in figure 2.5. Panels A, C, and E show tetrads of stimuli tested in the four preference-testing conditions and panels B, D, and F show the corresponding RWA visualizations and predictions. The first example neuron (figure 2.5A) responded to straight shafts tilted toward the upper right. This exemplifies the strict tuning for 3D properties of the solid shape, because even though the cell is tuned for a limb which lies in the fronto-parallel plane, the response is considerably higher for the solid versions of the stimuli (SP=0.624). The second example neuron (figure 2.5C) responded to curved shafts with a vertical orientation. This tuning to curved shafts could be parallel to tuning for orientation and curvature of contour fragments observed in earlier studies in the lab (Pasupathy and Connor, 2001). The third example neuron (figure 2.5E) responded to medial axis junctions positioned near the bottom right and opening toward the left. This neuron exemplifies tuning for junctions, a coding motif ubiquitous in the neural population. This, too, is reminiscent of earlier studies, particularly regarding the biased representation of sharp convex and concave curvatures in V4 (Carlson et al., 2011).

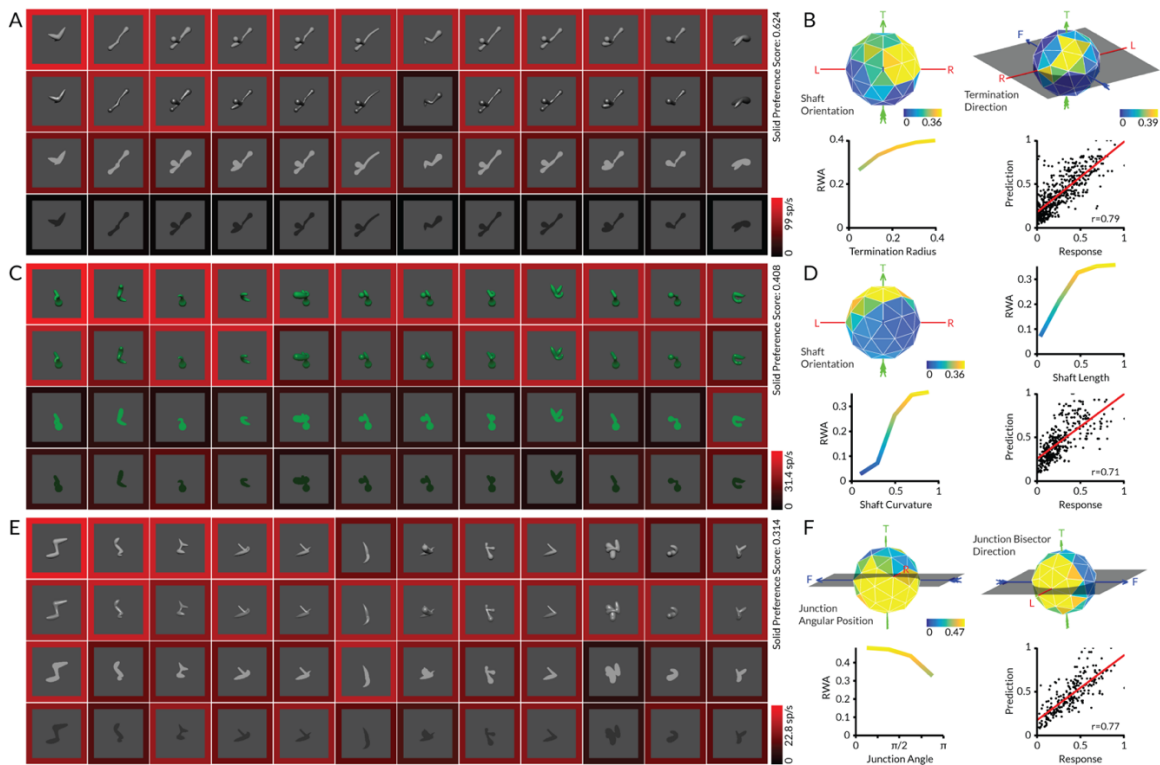


Figure 2.5: Examples of solid shape-prefering cells with tuning for solid shape fragments.

(A-F) Top tetrads for example V4 neurons depicting tuning for straight oriented shafts (A), curved shafts (C), and junctions (E). Slices through the tuning peak indicated with icosahedral and line plots. Prediction of neural responses based on lineage-product RWA in bottom right.

These examples are representative of tuning for terminations, straight/curved shafts, and junctions found in V4. Moreover, like these examples, the distribution of SP was significantly shifted away from 0 ($p = 0.0005$; t-test) with approximately two-thirds of the neural population (94/143 cells) preferred solid shapes and had an $SP > 0$ (figure 2.6A). Approximately 22% of the neurons (32/143 cells) had an $SP > 0.33$, which corresponds to the average solid shape response being at least 1.5 times greater than the planar response. A fraction of neurons responded more strongly to planar shapes (examples shown in figure 2.10 at the end of the results section), but, qualitatively, they tended to be neurons that were tuned to darker contrasts and/or very small stimuli. Figure 2.6B shows the distribution of predictions of neural responses with the RWA matrices; high

prediction accuracies indicate good convergence across lineages and effective parameterization of shape space. Across the three categories, prediction accuracies had a median greater than 0.65 and were significantly shifted from 0 ($p < 10^{-106}$). Prediction accuracies tended to be greater for shaft RWAs versus termination and junction RWAs (figure 2.6C-E) ($p < 0.005$; Kolmogorov-Smirnov test).

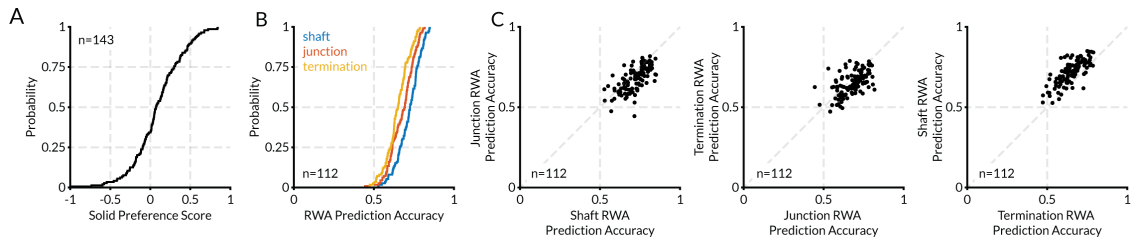


Figure 2.6: Population analysis of SP and RWA.

(A) Cumulative distribution of solid shape preference index for neurons with two or more generations. (B) RWA prediction for shaft, junction, and termination RWA matrices for all neurons with three or more generations. (C) Comparisons of prediction accuracies of shaft, junction, and termination RWAs. See supp. fig. 2 for examples of 2D selective cells.

2.2.3 Robust representation across 3D cues

On some experimental sessions, after the adaptive algorithm produced converging evidence of neural tuning from the two independent lineages, several post-hoc tests were performed. These tests served to ensure that the observed solid shape selectivity was not explained by changes in (a) contrast gradient across the object surface, (b) surface textural cues or other image cues, or (c) interaction between the self-occlusion boundary and surface contrast. These tests were run on four or more stimuli from each lineage, sampled from the low, medium, and high response ranges. This was done to ensure that the relative responses to solid shapes are preserved across varying cues.

Contrast post-hoc test

The contrasts of the solid and planar shapes were systematically varied as depicted in figure 2.7A. The responses to a high response shape are shown for four neurons. While the neurons are maximally responsive for a certain range of contrast, the difference between solid and planar shape responses remains high throughout the range. The SP was calculated for the stimuli tested in the contrast test and was compared with the SP from the adaptive algorithm run prior. It was found to be consistent across 15 tested cells (figure 2.7B) (r -squared = 0.27, p = 0.045 vs constant model). Because the solid shape stimuli in this test lacked specular highlights, three neurons that preferred specular shapes to shaded shapes were excluded from this analysis.

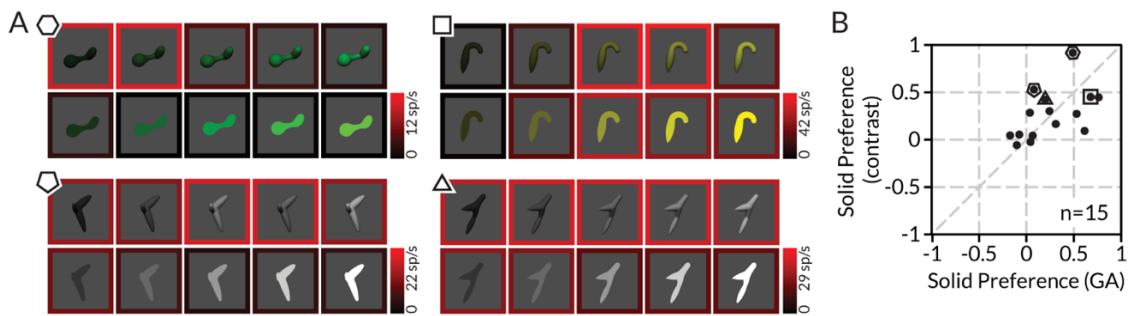


Figure 2.7: Contrast post-hoc test.

(A) Solid matte stimuli in increasing order of brightness and planar stimuli with increasing contrast. Responses are indicated with red image borders (see scale bars for individual neurons). *(B)* Comparison of solid shape preference index calculated with shapes from the adaptive shape experiment and from the contrast post-hoc test. Polygon markers correspond to neurons in (A).

Naturalistic images post-hoc test

To test if the tuning for solid shape was maintained if the shading and specular cues were replaced with other cues, the selected post-hoc objects were rendered with the ray tracing software Blender, with reflective or refractive optical properties. These objects, when placed in a naturalistic environment, reflected or refracted the features on the floor and/or walls to produce a striking percept of solid shape structure. Importantly, the images produced by this method were

entirely different from the shading-based stimuli on a pixel-by-pixel basis, similar only in their inferred shape structure. Figure 2.8A shows responses of seven neurons to example top and bottom stimuli rendered in naturalistic conditions. Here, “top” and “bottom” do not refer to the highest and lowest responses in this test, but to example high and low response stimuli selected from the adaptive algorithm. Not only were the neural responses found to be consistent across rendering conditions, the disparity between top and bottom stimulus responses was consistent between this test and the adaptive algorithm (figure 2.8B). The difference between the top and bottom responses was significantly greater than 0 ($p = 0.011$; t-test).

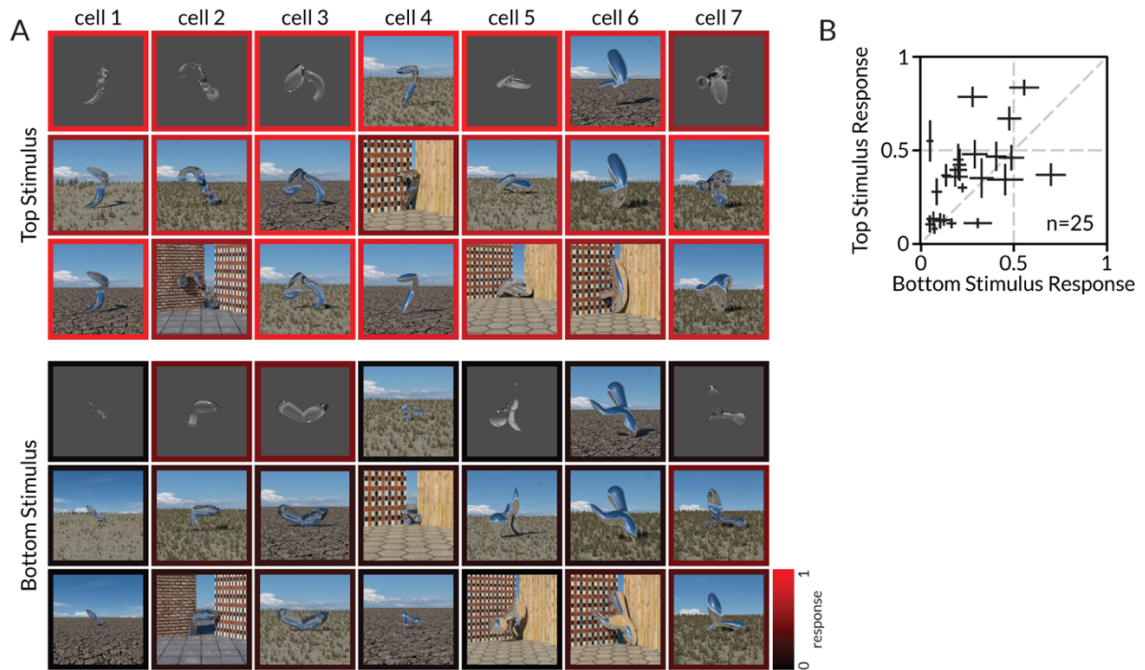


Figure 2.8: Naturalistic images post-hoc test.

(A) High and low response stimuli (top and bottom) from the adaptive algorithm rendered in naturalistic conditions also have high and low responses. (see 2.2.11 at the end of the chapter for the full set of stimulus conditions.) (B) Neural responses to naturalistic shape stimuli as a proportion of their responses in the adaptive algorithm.

Stereogram post-hoc

As a final test for solid shape preference, the solid and planar shapes were rendered using random dot stereograms (RDSs). RDS stimuli lack any image-level or monocular cues and rely purely on the binocular disparity between dots projected to either eye to create a percept of depth and solid shape. Both solid and planar stereograms were tested at 3-5 depths behind, at, and in front of the fixation plane. A field of background dots was placed behind the farthest depth to provide depth context and remove the ability to perceive the shape silhouette. Figure 2.9 shows responses of four example neurons to solid and planar stereograms at three depths. The neural responses are consistently higher for the solid stereogram at the far and fixation-plane depths. The SP, now calculated for the stimuli in this test alone, shows a similar distribution of solid shape preference as the distribution obtained from the adaptive algorithm (compare figure 2.9B with 2.6A). This distribution is significantly shifted from 0 ($p = 0.0098$; t-test). This test definitively demonstrates that V4 solid shape selectivity cannot be explained by image-level features like contrast gradients, pixel-by-pixel correlations, 2D contour curvature, etc.

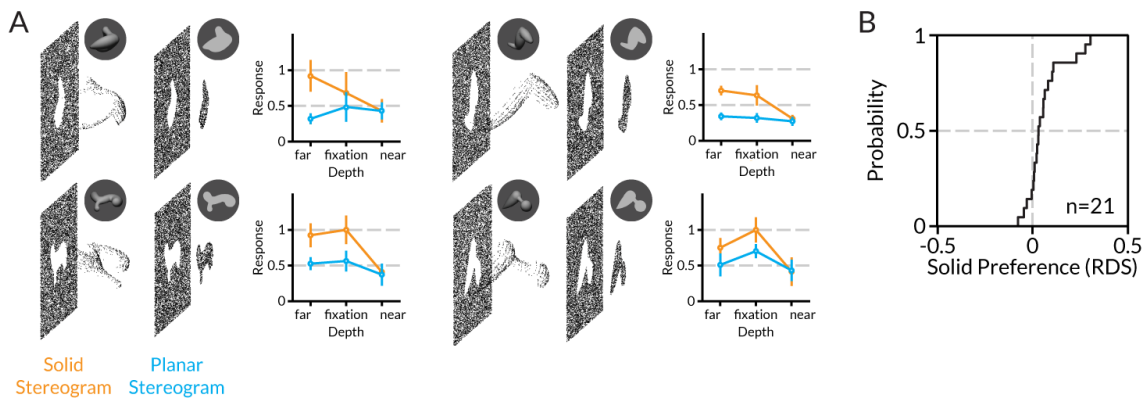


Figure 2.9: Stereogram post-hoc test.

(A) Neural responses to solid and planar stereograms of high response shapes for four neurons presented at three depths relative to the fixation plane. Background dots are presented at a constant depth behind all shapes. (see figure 2.12 at the end of the chapter for example stereograms). (B) Cumulative distribution of solid shape preference scores calculated only using the stereogram post-hoc test.

2.3 Discussion

The results from this chapter address fundamental questions about the transformation from 2D image signals to solid shape perception. Previously, intermediate transformations in object vision have been studied with planar image features, such as contour fragments and Gabor patches. In the experiments described in this chapter, microelectrode recordings in awake, fixating macaque monkeys were used to study the responses of neurons in area V4 to complex solid shapes, and to compare them to responses evoked by the shapes' planar counterparts. Random solid shapes were created by constructing a medial axial structure out of limbs with randomized connectivity, orientations and curvature, each wrapped with a tubular surface of a random radius profile. Neurons in V4 were reliably tuned for the orientation and curvature of solid shape fragments and object-relative position. Crucially, these tuning properties were consistent across 3D cues like shading, disparity, reflectivity, and refraction. Images rendered with these cues cohere only in their implicit mapping to solid shape and invariant responses to such stimuli can only be explained by an underlying representation for 3D geometry. Response-weighted averaged (RWA) matrices with dimensions for 3D geometric fragments like the width and curvature of shafts and junctions, predicted the neural responses with high accuracy. Current models of object vision rely on image transformations for information about object identity. These results now introduce the extraction of solid shape geometry by early primate vision as an alternate path to object recognition.

2.3.1 Solid shape coding unifies contour and spectral coding hypotheses

Previous studies concentrated on coding for planar image fragments and spectral properties in V4 neurons. More than three decades ago, V4 was found to be critical for form vision (Desimone and Schein, 1987) in addition to its already established contribution to color vision (Zeki, 1983, 1973). Building on these observations, (Kobatake and Tanaka, 1994) studied V4 by sampling a

large stimulus set of objects, faces, bars, gratings, colored contours etc. to show that V4 neurons preferred stimuli with intermediate complexity compared to those preferred by V2 or TE neurons. Similar results were obtained by (Gallant et al., 1993, 1996), who used a parameterized stimulus set of Cartesian, hyperbolic, and polar grating patterns to probe V4 neurons. By demonstrating tuning to hyperbolic and polar gratings in a significant fraction of V4 neurons, this study provided clear evidence for tuning for a more complex property than observed in earlier visual areas. Further refinement of the understanding of V4 tuning functions came from a later set of studies that parametrically changed, among other things, the orientation and curvature of shape contour fragments, as well as their object-relative position (Pasupathy and Connor, 1999, 2001). V4 neurons were found to be tuned to all of these features, which explained and extended the findings of the previous studies. Similarly, V4 responses were also found to be tuned to the orientation of disparity-defined edges (Hegd  and Van Essen, 2005; Shiozaki et al., 2012; Umeda et al., 2007) and 3D orientations (Hinkle and Connor, 2002). Generally, the observed V4 response properties are assumed to arise by calculating a weighted sum of inputs from orientation- and disparity-tuned neurons in V1. In summary, these studies have led to the overarching hypothesis that V4 neurons are tuned to a part of an object (Connor et al., 2007).

An alternate theory of V4 function posits that its neurons are not feature detectors or tuned for object parts, but instead are spatiotemporal filters which calculate a spectral transform on the input image (David et al., 2006; Touryan and Mazer, 2015). Under this theory, all neurons from V1 to V4 (and perhaps even beyond) essentially calculate the same transformations on their input. The sequential increase in complexity of the representation is then simply a reflection of these repeated calculations. Indeed, researchers have attempted to model neural tuning to shape properties in V4 with spectral receptive fields, but these studies have been only partially successful (Oleskiw et al., 2014). However, the same mechanistic theory has inspired deep

convolutional network architectures for object recognition, which are currently the most successful attempts at computer vision.

The results in this chapter provide a possible route for reconciling these two competing hypotheses. First, the solid shape tuning reported in this chapter is directly related to the tuning for planar orientation and curvature observed before. The RWA analysis shows that neurons have tuning peaks in a space defined by, among other things, the orientation and curvature of curved shafts. This tuning was also found to be sensitive to the object-relative position of the shape feature, again recapitulating earlier results. Second, tuning for 3D orientations and disparity-defined edges is also observed in the results in this chapter. The results from the stereogram post-hoc show that the neurons are tuned to solid shape-in-depth rendered with disparity cues alone. Also, the preponderance of tuning for shafts oriented in 3D parallels the tuning for 3D orientations found previously in V4. Therefore, the results in this chapter reproduce previous findings of planar contour and disparity tuning in V4 and offer a unifying theory of neural tuning based on solid shape representations.

At the same time, these results are also compatible with the spectral filter theory of V4, in which neural tuning is explained by responses to spatial frequency patterns. Spatial frequency patterns create a perceptual blur effect reminiscent of solid shape shading. Tuning for spatial frequency patterns in V4 could therefore possibly be explained by tuning for solid shape fragments and their underlying lighting patterns. Indeed, the joint coding of contour and blur has been studied recently in V4 (Oleskiw et al., 2018). While a systematic, comparative study of the lighting and blur patterns on solid shapes that create various spatial frequencies would be a vital confirmation, solid shape representation in V4 is the first evidence-backed theory of V4 tuning that attempts to bridge the gap between the competing hypotheses of contour-based and spatial-frequency-based models of V4.

2.3.2 The need for and utility of solid shape processing in V4

The prevailing dogma of solid shape representation, as discussed in chapter 1, is that all coding transformations before TE are in the image plane. This includes the assumption that V4 neurons represent planar contour or spatial frequency features of a given image, a representation that is later integrated to give rise to solid shape signals in TE. The results in this chapter challenge this dogma by showing explicit solid shape representation as early as V4. These results fundamentally alter current theories of coding transformations in V4, and by extension, the ventral visual pathway.

Information about solid shapes is critical to object perception and making this information explicit for downstream areas that guide movements, assign value, and make decisions based on object geometry is critical for efficient transfer and decoding. Furthermore, coding for 3D reality is more intuitive and explicit than image transformations, is more robust to changes in viewing conditions and occlusion than planar image transformations and covers a considerably larger representational space.

This information was thought to be extracted in the later ventral stream areas like TEO and TE. The result that solid shape information is made explicit as early as V4 fundamentally challenges that theory. TEO and TE are major targets of projections from V4. Neurons in TEO begin to represent combinations of shape parts to complex configurations to form a cohesive object representation (Brincat and Connor, 2004). Neurons in TE encodes complex solid shapes as combinations of medial axial components and surface fragments (Hung et al., 2012). The origin of these complex properties was previously unclear but the results from the current study would suggest that TEO neurons non-linearly combine surface fragments encoded by single neurons in V4. Indeed, the results described in this chapter could be construed as the source of the solid medial axis and surface shape signals that would be non-linearly combined to form the known TE

representation for solid shapes. Feedforward projections from V4 also terminate in ventral posterior parietal regions, and regions in the intraparietal sulcus like LIP, VIP, and CIP. The latter regions are known to be involved with saccadic eye movements, visually guided reaching, and perception of depth from stereopsis. The presence of solid shape responses in V4 implies that visually guided behavior must be influenced by these early solid shape signals.

The consistency of solid shape responses across multiple 3D cues as early as V4 challenges previously held notions of invariant coding. This consistency is certainly beyond what is possible with image transformations alone and raises interesting questions about the origin of these responses and the inputs on which these transformations occur. Do these responses originate earlier than V4 in texture responses in V1 and V2? Or are orientation-, disparity-, and texture-tuned inputs integrated by V4 neurons to give rise to these invariances? Or are these responses a reflection of recurrent connections within V4 or feedback connections from IT cortex? Future investigations into these questions will provide key insights into ventral visual pathway transformations.

In conclusion, the results in this chapter demonstrate that solid shape is encoded as early as V4 in the ventral visual pathway. They also resolve conflicting hypotheses of disparate tuning properties in V4 neurons by introducing a superset of higher-dimensional coding transformations based on solid geometry. These results fundamentally challenge the current view of coding transformations in the ventral visual pathway (see general discussion in chapter 4 for more details).

2.4 Methods

2.4.1 Experimental Model and Subject Details

Two adult male rhesus macaques (*Macaca mulatta*) weighing 8.0 and 11.0 kg were used for the awake, fixating, electrophysiology experiments. They were singly housed during training and experiments. All procedures were approved by the Johns Hopkins Animal Care and Use Committee and conformed to US National Institutes of Health and US Department of Agriculture guidelines.

2.4.2 Method Details

Behavioral task

Both monkeys were head-restrained and trained to maintain fixation within a 0.5deg window (radius) surrounding a 0.25deg square sprite (fixation spot) displayed on a monitor 60cm away for juice reward. Eye positions for both eyes were monitored with a dual-camera, infra-red eye tracker (ISCAN, Inc, Woburn, MA). In trials requiring stereoscopic fusion, separate images were shown in left and right eyes via cold mirrors. In trials not requiring stereoscopic fusion, the same images were shown. Stereo fusion was monitored with a random-dot stereogram search task.

Neural recording

The electrical activity of 169 (114 and 55 respectively from the two monkeys) well-isolated neurons was recorded using epoxy-coated tungsten electrodes (FHC Microsystems), processed with TDT RX5 Amplifier (TDT, Inc, Alachua, FL). Single electrodes were lowered through a metal guide tube into dorsal V4, targeted with a custom-built electrode drive. Area V4 was identified on the basis of structural MRI, the sequence of sulci as the electrode was lowered, and the visual response characteristics of the neurons. The neurons receptive field properties (size,

position, and color preference) were mapped using 2D sprites (bars and random planar shapes) under experimenter control.

Visual stimulus construction and morphing

Solid stimuli were constructed using a procedure similar to Hung et al (Hung et al., 2012) (see figure 2.2). Briefly, stimuli were generated by connecting 2-4 medial axial components. The skeletal structure – limb configuration, limb lengths, curvatures, and widths were randomly generated. The width profile of each limb was generated using a quadratic function fit to the randomly generated width at the mid-point and the two ends. Limb junctions were smoothed using a gaussian kernel on the 3D position and normal for every face. During the experiment, the stimuli were morphed by adding, deleting, or replacing limbs, or changing the length, orientation, curvature, or width of the limbs. Additionally, the position, size, and 3D orientation of the object were also changed probabilistically. Planar shapes were generated by turning off shading/lighting in OpenGL.

Adaptive stimulus algorithm

Each neuron was studied using two independent lineages of evolving solid shapes (see figure 2.1). The first generation consisted of 40 randomly generated solid stimuli (matte or shiny shading) in each lineage. Each subsequent generation was divided into two parts - shape evolution and solid shape preference testing. For shape evolution, 16 stimuli were randomly selected for morphing from the previous generation - 6 from the stimuli from the top 10% response range, 4 from the next 20%, 3 from the next 20%, 2 from the next 20%, and 1 from the bottom 10%. An additional 4 randomly generated stimuli were added to the pool. For solid/planar preference testing, the top 5 stimuli from the previous generation were rendered in solid matte, solid shiny, planar high contrast, and planar low contrast versions. The contrasts of the planar stimuli were chosen to be the average contrast across solid matte and shiny shapes. This

algorithm allowed us to not only to test each neuron's solid shape tuning based on its responses to the stimuli, but also simultaneously test its solid shape preference. Neurons were tested with 80-440 randomly generated and morphed stimuli. For the 143 neurons studied with two or more generations of the adaptive algorithm (102 of 114 and 41 of 55 neurons from the two monkeys respectively), the solid preference score was calculated as:

$$SP = \frac{R_{solid} - R_{planar}}{\max(R_{solid}, R_{planar})}$$

where R_{solid} is the average neural response across matte or polished solid stimuli, whichever is higher, and R_{planar} is the average neural response across high or low contrast planar stimuli, whichever is higher. This score varies between -1 (prefers only planar stimuli) to 0 (no preference) to +1 (prefers only solid stimuli).

Post-hoc stimulus selection and construction

After the evolutionary testing procedure, 4-8 stimuli were sampled from the high (top 20%), medium, and low (bottom 20%) response ranges to be tested with the following post-hoc tests (see figure 2.7 to 2.9).

Contrast: Each stimulus was rendered as a planar shape with 5 levels of increasing contrast (0.2, 0.4, 0.6, 0.8, 1 on a background of 0.3 gray) or a solid matte shape with 5 levels of increasing surface brightness (figure 2.7).

Naturalistic stimuli: Each selected stimulus was rendered with a smooth, purely reflective or purely refractive (index 2.0) shader in Blender. It was then placed in a closed (surrounded by walls and a floor) or in an open (field of grass or textured ground) environment. The textures were selected from online texture libraries. (figure 2.8)

Stereogram: The visible parts of each solid (or planar) shape were selected to generate a solid (or planar) stereogram by randomly placing dots at the appropriate depth. This shape was then translated in depth relative to the fixation spot. A set of random background dots were placed behind the shape (at the farthest depth). 3-5 relative disparities were tested for each neuron. (figure 2.9)

2.4.3 Response-weighted Averaging Analysis

The RWA analysis was performed independently on each lineage on neurons studied with more than 80 solid stimuli (more than three generations) per lineage.

For metric shape analysis, each shape was divided into its constituent limbs and each limb was described in terms of the geometry of its terminal end and its shaft (figure 2.2). Additionally, the junctions between pairs of limbs were also independently parameterized. These geometric constructs were parameterized using their position relative to the center of the shape, their orientation in 3D space, and their lengths, curvatures, and radii.

Based on these parameterizations, RWA matrices were constructed – one each for shafts, terminations, and junctions – such that each element of these matrices is a bin that represents a specific part of any shape (figure 2.3). The dimensions of each of these matrices are the geometrical parameters (figure 2.3) used to describe the shape fragment. For example, the termination matrix has four dimensions – 3D angular position (80 bins on a icosahedral surface), 3D radial position (5 bins), 3D direction (80 bins on a icosahedral surface), and radius (5 bins). To populate these matrices, for each shape, several response-weighted gaussians (centered at the bins occupied by every shape fragment in a solid shape) were iteratively summed into the matrices (figure 2.4). After repeating this for every shape, each matrix was normalized with a non-response-weighted matrix. Finally, the RWA matrices obtained from each independent

lineage were multiplied bin-by-bin to produce an accurate representation of neural tuning. These lineage-product RWA matrices were used in all further analyses and visualizations.

Each RWA matrix was used to predict responses to the full set of solid shape stimuli presented during the experiment. Because the responses of most cells were well predicted by all three RWA matrices (figure 2.6), a linear combination of these predictions was fit to obtain the final prediction of neural responses.

3 Micro-organization of solid shape representation

3.1 Introduction

The horizontal organization of specific tuning dimensions on the cortical sheet is considered to be an important indication of the representational space in an area (Kaas, 1997) (also see section 1.4.2). Given that most computations are performed locally, the functional organization of an area can shed light on how the coding transformations in it might be implemented. The layout of tuning dimensions likely underlies the reduction of dimensionality of representational space onto the cortical surface by optimizing axonal wiring length (Allman and Kaas, 1974; Chorniak, 1994; Chklovskii and Koulakov, 2000). While the utility of cortical organization is a matter of debate (Horton and Adams, 2005; Nauhaus and Nielsen, 2014), its study nonetheless can inform models of inter-areal addressability, recurrent computations, and temporal dynamics of neural responses. Cortical organization can be understood on many spatial scales: fMRI has been used to reveal clustering of semantic concepts (Huth et al., 2016), face patches (Tsao et al., 2003, 2006), color modules (Conway et al., 2007), task-related activity (Glasser et al., 2016) etc., optical imaging has been used to reveal orientation and ocular dominance columns in V1 (Blasdel and Salama, 1986; Bonhoeffer and Grinvald, 1991), and two-photon imaging has been used to reveal the mutual organization of orientation with spatial frequency, hue, and ocular dominance on a cellular level in V1 (Garg et al., 2019; Nauhaus et al., 2012). Although clustering of tuning dimensions may be visible on larger spatial scales, making precise predictions of coding transformations requires an understanding of *micro-organization*, or neuron-level organization of an area. At larger spatial scales, cortical organization probably reflects average activations of large cortical ensembles, which can mask possible discrepancies and distortions that contribute to the richness of local coding transformations. Also, only high-resolution imaging techniques, like two-photon imaging, can definitely establish a lack of organization. The constraints on the micro-organization of higher-level visual areas, like V4, are largely unknown, and hypotheses of how

tuning dimensions emerge in these areas are yet to be scrutinized. Given the results from the previous chapter, knowing such constraints would empower theories of how solid shape representation emerges in V4. This chapter investigates the relative locations of solid or planar shape-preferring neurons and their respective tuning for solid/planar shape fragments. As with the findings in the previous chapter, a substantial fraction of neurons responded selectively to solid shapes versus planar shapes. Additionally, strong local clustering of solid shape preference in separable cortical patches was found and, importantly, these patches of solid/planar shape selectivity had tuning for congruent shape features. This organization of solid shape selectivity is the first demonstration of the functional micro-organization in V4 and facilitates hypotheses about the visual representational space in mid-level ventral visual areas.

3.2 Results

3.2.1 Neural responses to solid shapes with two-photon imaging

The calcium indicator Oregon Green BAPTA (OGB) dye was bulk-loaded into small durotomies in area V4 in three anesthetized macaque monkeys to investigate the organization of solid shape tuning. The injected region was imaged with a two-photon microscope about an hour after the injection. Imaging regions were about $500\mu\text{m} \times 500\mu\text{m}$ in size, at depths from the cortical surface ranging from $100\mu\text{m}$ to $220\mu\text{m}$. These depths correspond to upper and middle layer 2/3 of cortex. Neurons were differentiated from glial cells on the basis of selective uptake of sulforhodamine 101 (SR101) in glia, which was visible in the red channel (figure 3.1).

Each experiment started with mapping the aggregate receptive field (RF) of the imaging region by flashing planar shapes containing a range of curvatures on parts of the screen. These 2D stimuli were designed to have sharp and smooth convexities and concavities at various positions relative to the object center and were effective in eliciting responses for RF mapping (figure 3.2).

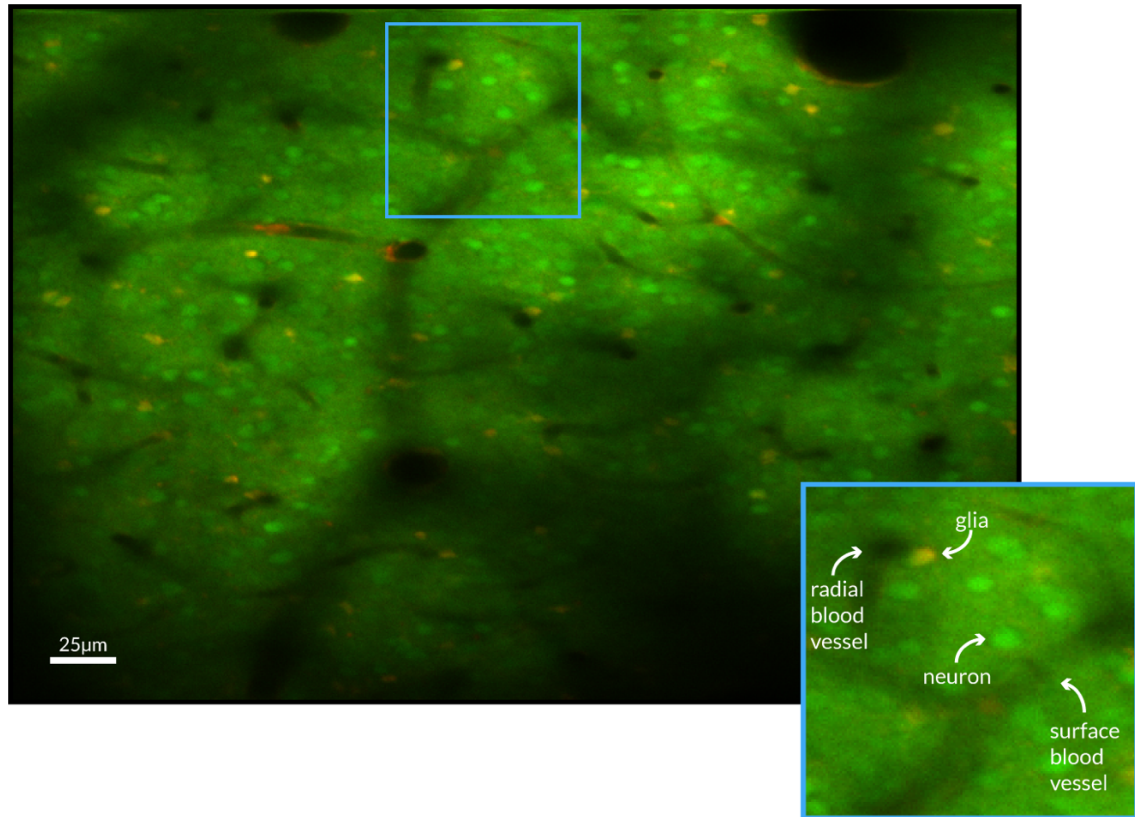


Figure 3.1: Example two-photon imaging region 1.

An approximately 300µm x 300µm region loaded with OGB+SR101 imaged with a two-photon microscope. Inset depicts a zoomed in version of the indicated sub-region. Neurons (green blobs) and glia (yellow blobs) are differentiated based on selective SR101 uptake by glia cells. Radial and surface blood vessels help identify and manually recalibrate the microscope position in case there is considerable drift during the experiment.

Mapping stimuli varied in size and rotation at each position on the screen (figure 3.2A) and were arranged in horizontal and vertical arrays to efficiently sample the visual field (figure 3.2B). The position that elicited the maximum activation across the imaging region was selected to be the aggregate RF for the region. In some cases, the RF was further refined by confining the mapping experiment to a smaller area centered at the previously estimated RF.

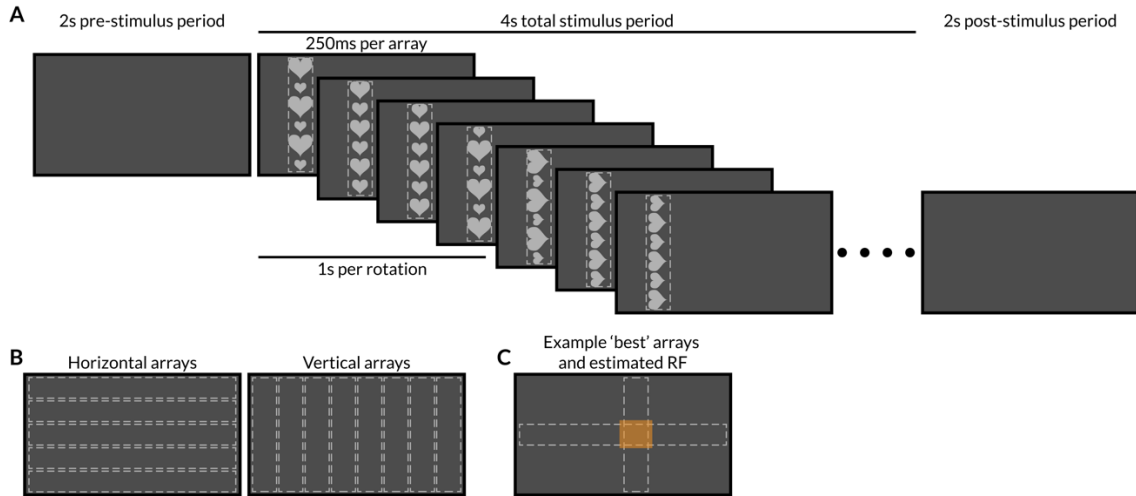


Figure 3.2: Receptive field mapping during two-photon imaging.

(A) A schematic of a single trial in the receptive field mapping experiment. 2D heart-shaped stimuli are displayed in a horizontal or vertical array at varying sizes and orientations. Each array is flashed for 250ms, and each location is sampled with four sizes and four rotations of the stimuli in a single trial. The neural responses are averaged for every trial and across repetitions for a given array. **(B)** The locations of horizontal and vertical arrays. **(C)** The intersection of the horizontal and vertical array locations that elicit the highest responses is taken to be the aggregate receptive field of the imaging region.

Imaging time per region was limited in two-photon experiments because of photobleaching and image degradation due to tissue growth. An adaptive sampling experiment (as described in chapter 2) was therefore not feasible. Instead, two types of conventional shape mapping experiments were devised with sets of simple solid and planar stimuli, chosen to span the surface curvature and 3D orientation space (figures 3.3 and 3.4). The first set consisted of C- and L-shaped stimuli, each sized to fit the estimated RF. For this set, each stimulus was flashed at the position of the RF for 1-2s, with an inter-stimulus interval of 2-4s (figure 3.3). The second set consisted of large (full-field), tubular, L-shaped stimuli. More precisely, each stimulus consisted of repetitions of the same L-shaped stimulus, positioned so that between them they formed a pattern that covered the entire screen width. Stimuli were drifted for two full cycles of the pattern

(figure 3.4). As with the previous set, these shapes effectively spanned the relevant surface curvature and 3D orientation space. However, in contrast to the first stimulus set this second set offers the advantage of being independent of the RF location measurement. To further achieve this goal, stimuli were displaced in small steps in the direction orthogonal to the drifting direction across trials. Stimulus position ranged from one edge of the screen to the other to ensure that each shape fragment was drifted within the RF of every neuron (figure 3.4B). The results from both sets of experiments were qualitatively similar and therefore grouped for the rest of the analyses.

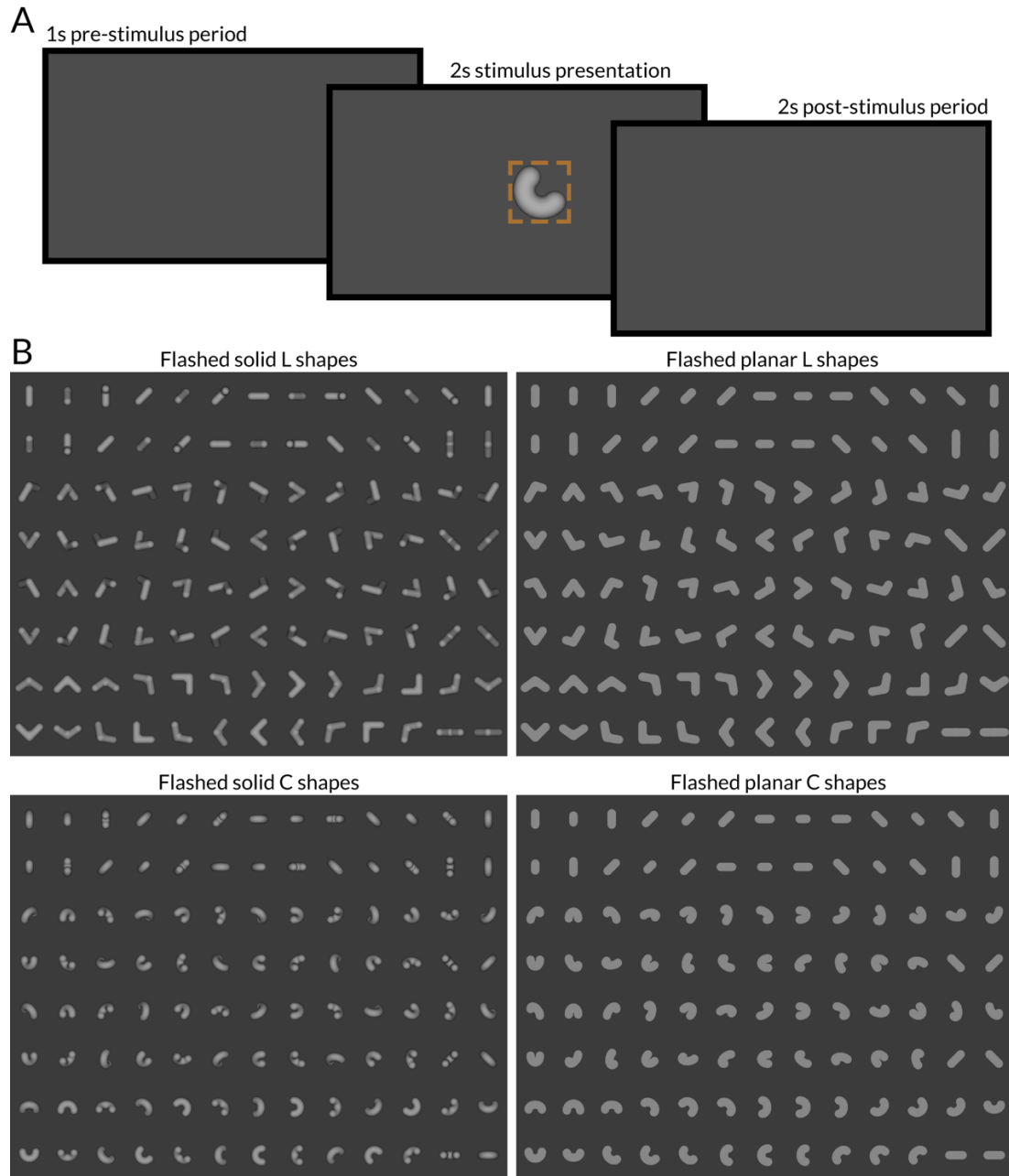


Figure 3.3: Trial structure and stimulus set 1 used in two-photon imaging experiments.

(A) A schematic of a single trial in the solid shape mapping experiment with stimulus set 1. A single stimulus from a coarsely sampled set of stimuli from **(B)** is presented within the estimated aggregate RF. **(B)** L- and C-shaped solid and planar stimuli rotated in 45° steps. Some planar counterparts are identical and are therefore presented multiple times during the experiment.

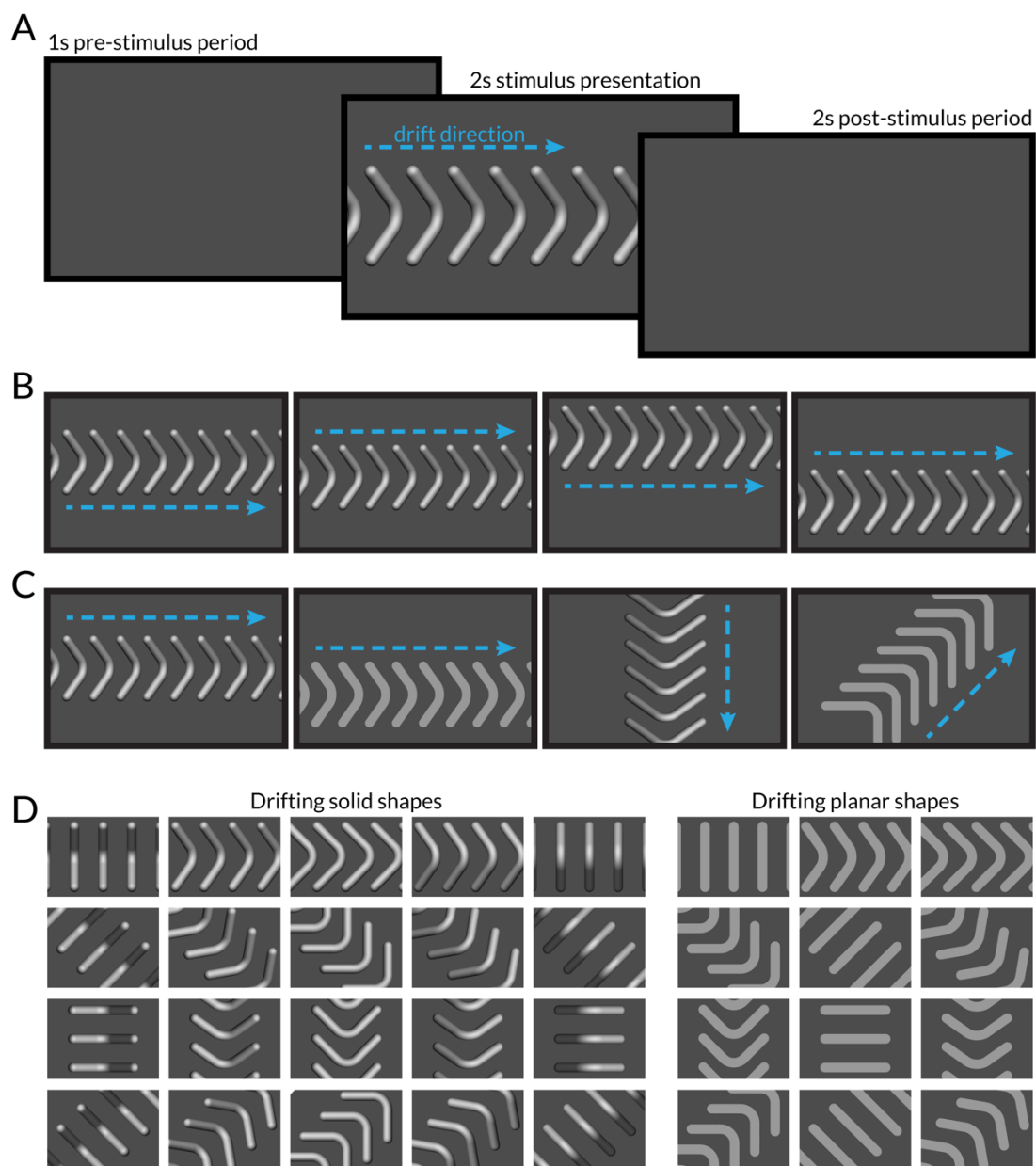


Figure 3.4: Trial structure and stimulus set 2 used in two-photon imaging experiments.

(A) A schematic of a single trial in the solid shape mapping experiment with stimulus set 2. A series of oriented bent tubular shapes was drifted for two cycles on the screen. The displacement orthogonal to the drift direction is randomly chosen on any given trial out of 11 positions spanning the entire screen.

(B) Four example trials showing a subset of displacements sampled for a single stimulus in this experiment.

(C) Four example trials showing a subset of solid and flat shape conditions and different drift directions.

(D) The full set of stimuli sampled at 45° intervals of in-plane and out-of-plane rotations.

After each experiment, the images obtained were corrected for motion artifacts in the horizontal plane (see methods section 3.4.2), and a binary mask was manually applied to define the set of pixels belonging to a neuron. Glia were removed from further analyses at this point. The fluorescence time-course ($\Delta F/F$) of each neuron for each stimulus trial was calculated as $\frac{\Delta F}{F_0} = \frac{F(t) - F_0}{F_0}$ where $t=0$ represents the stimulus onset, $F(t)$ is the pixel-averaged fluorescence response, and F_0 is the pixel-averaged fluorescence averaged across the baseline period 1200-200ms immediately before stimulus onset. Examples of the fluorescence time-courses are depicted in figure 3.5. The fluorescence sharply rises between approximately 50-150ms after stimulus onset and decays slowly to baseline levels by approximately 800ms after stimulus offset. Therefore, the stimulus response for each neuron is calculated as the average $\Delta F/F$ between 200ms after stimulus onset and 800ms after stimulus offset. In this case, only neurons that passed a 2D Kolmogorov-Smirnov test (Justel et al., 1997) between the top- and bottom- stimulus were analyzed (details in the methods section 3.4).

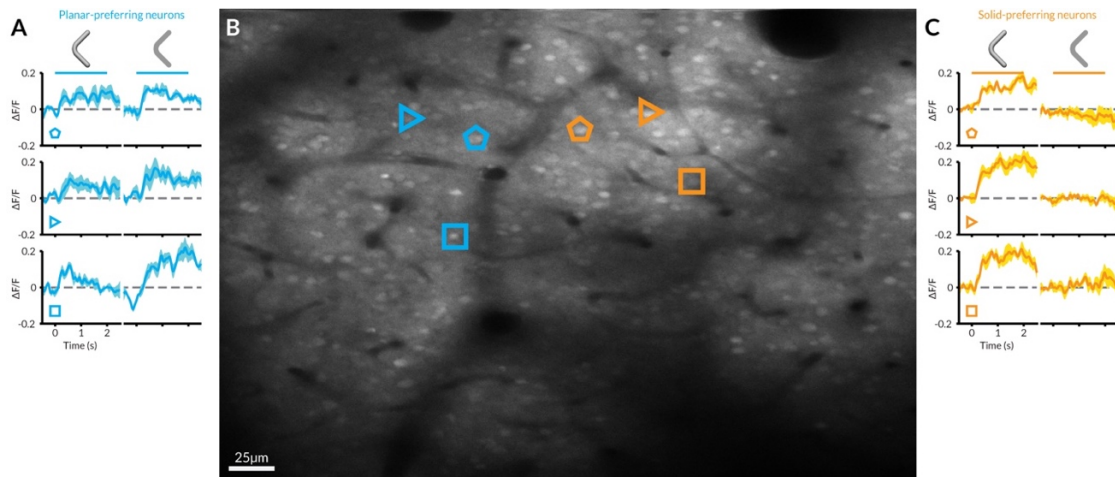


Figure 3.5: Fluorescence time course for solid and planar shape-prefering neurons.

(A) Peri-stimulus time-courses for three planar stimulus preferring neurons. Horizontal bars signify the stimulus presentation duration (2s) of the stimulus indicated by the icon above it. Polygon icons indicate the location of the neuron in the imaging region in (B). Solid lines are the trial-averaged time-course and translucent fill indicates the SEM. (B) Average green-channel image for the same region as in figure 3.1. Polygon icons indicate the neurons in (A) and (C). Scale bar is shown on the bottom left. (C) Same as (A) for solid shape-prefering neurons.

3.2.2 Organization of solid shape preference

For the example imaging region depicted in figure 3.5, the time courses suggested that at least a subset of neurons have a strong preference for solid shapes versus the corresponding planar shapes. This was quantified for all neurons in the imaging region using the same formula for solid preference (SP) as the one used in the previous chapter.

$$SP = \frac{R_{solid} - R_{planar}}{\max(R_{solid}, R_{planar})}$$

where R_{solid} is the average neural response across solid stimuli and R_{planar} is the average neural response across planar stimuli. This score varies between -1 (prefers only planar stimuli) to 0 (no preference) to +1 (prefers only solid stimuli).

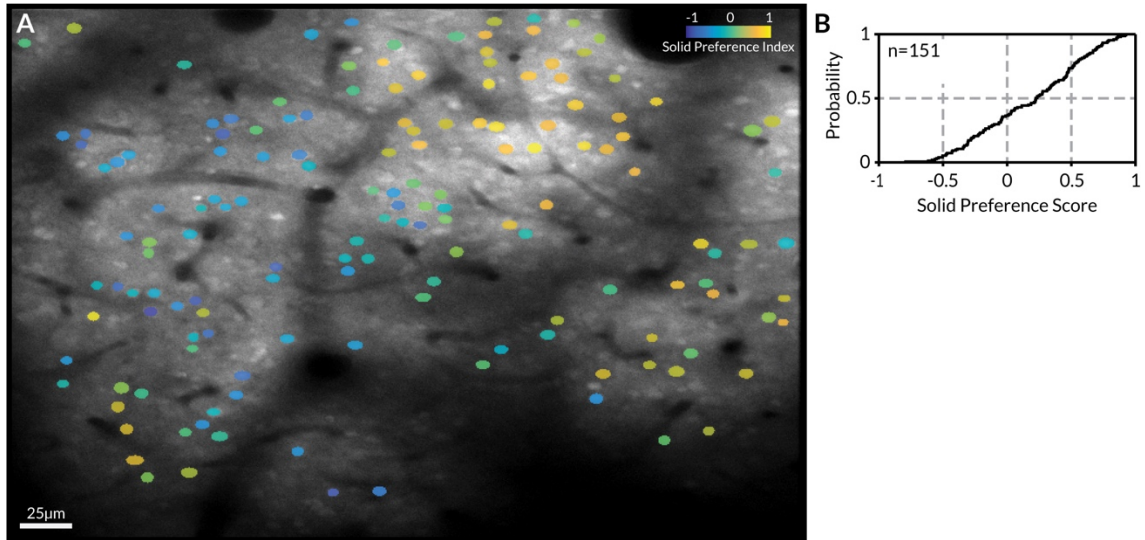


Figure 3.6: An example functional map of solid shape preference in V4.

(A) SP index overlaid on the anatomical average image (same as figure 3.1, 3.3). Only neurons that pass the significance test (see methods) are filled in. Color bar for SP index in the top right. **(B)** Distribution of SP values from the region in (A).

This analysis revealed two insights. First, it demonstrated that the distribution of SP values for this region showed a significant bias towards positive SP values (figure 3.6B) (t-test; $p < 10^{-6}$). Second and more importantly, the map of SP values across the imaging region (determined by color coding each neuron with its SP score) revealed a strong clustering of solid shape-preferring neurons towards the right of the region (perhaps extending further to the right) and another region in the bottom-left (perhaps extending further to the left) and a cluster of planar shape-preferring neurons in the middle of the region (figure 3.6A).

Further, figure 3.6A suggested that the preference for solid versus planar shapes changes rapidly (rather than gradually) between the different clusters. To quantify the rate of SP change across cluster boundaries, the boundaries between adjacent clusters was marked with a novel spatial clustering technique. In addition to labeling the cell soma, OGB also labels the interstitial space

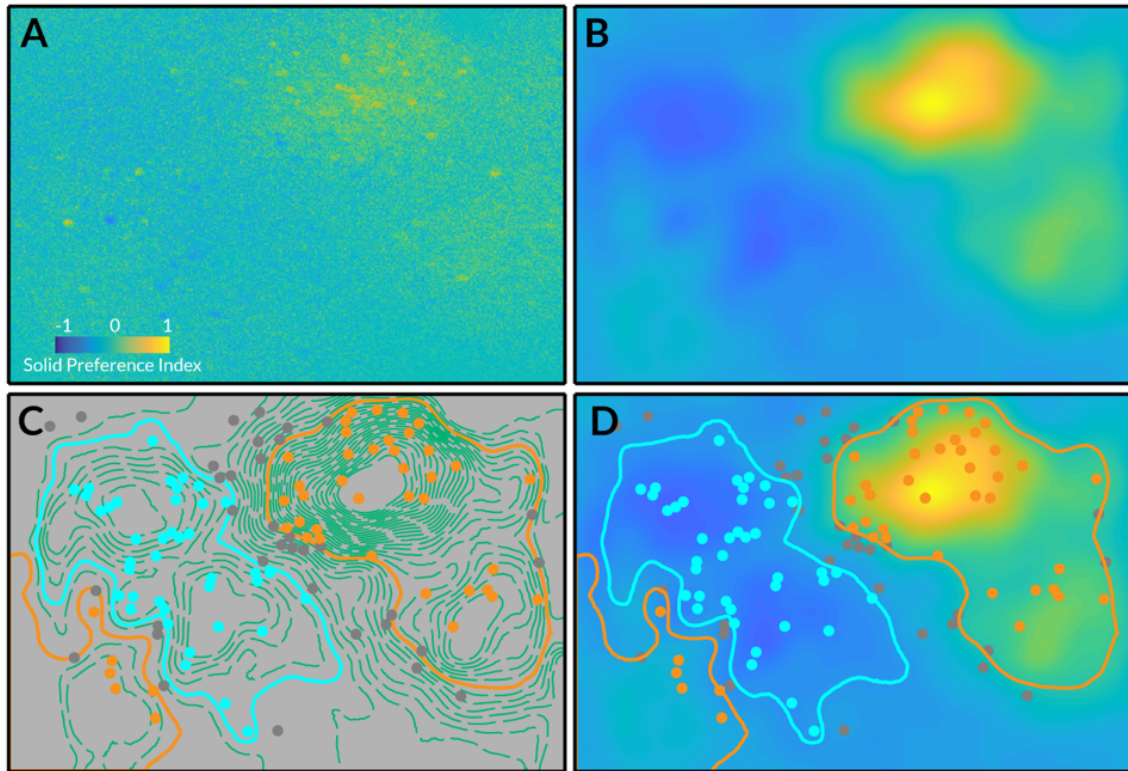


Figure 3.7: Spatial clustering analysis based on pixel-SP functional map.

(A) Map of SP values for each pixel for the imaging region in figure 3.1. (B) The same map smoothed with a Gaussian kernel with broad width. (C) Contour map (green dashed lines) of the smoothed map in (B). Dots indicate neuron locations. Cyan and orange contours are the contours selected as cluster boundaries. (D) Selected contour boundaries and neuron locations overlaid on smoothed pixel-map from (B).

between cells (neuropil). Furthermore, the neuropil generally appears to be tuned to similar stimulus properties as the embedded neurons (see Figure 3.7 for an example), albeit with a much weaker signal strength. Because the neuropil signal is spatially continuous throughout the imaging region, it can more easily be used to generate a continuous map for the imaging region, rather than the punctate signal generated by cell somas depicted in figure 3.6. To generate the

neuropil map, a peri-stimulus time-course was extracted for each pixel (instead of every neuron). Based on these responses, an SP value was assigned to each pixel creating a continuous map of SP (figure 3.7A). This map was then smoothed and normalized (figure 3.7B). This smooth map produced iso-preference contours that were considered as candidates for clusters (figure 3.7C). These cluster contour candidates were then short-listed based on criteria defined in the methods section 3.4.3. Briefly, contours that contained fewer than four neurons and those that were smaller than 10% of the size of the imaging region or were too complex (based on the ratio of the contour area and the area of its convex hull) were rejected. Of the remaining contours, the largest and most homogeneous contours were selected based on their area and the kurtosis of the distribution of SP values of the neurons contained in the contour. The final set of cluster contours were smoothed using a spline smoothing function. The cluster contours for the example region from figure 3.6 are shown in figure 3.8A.

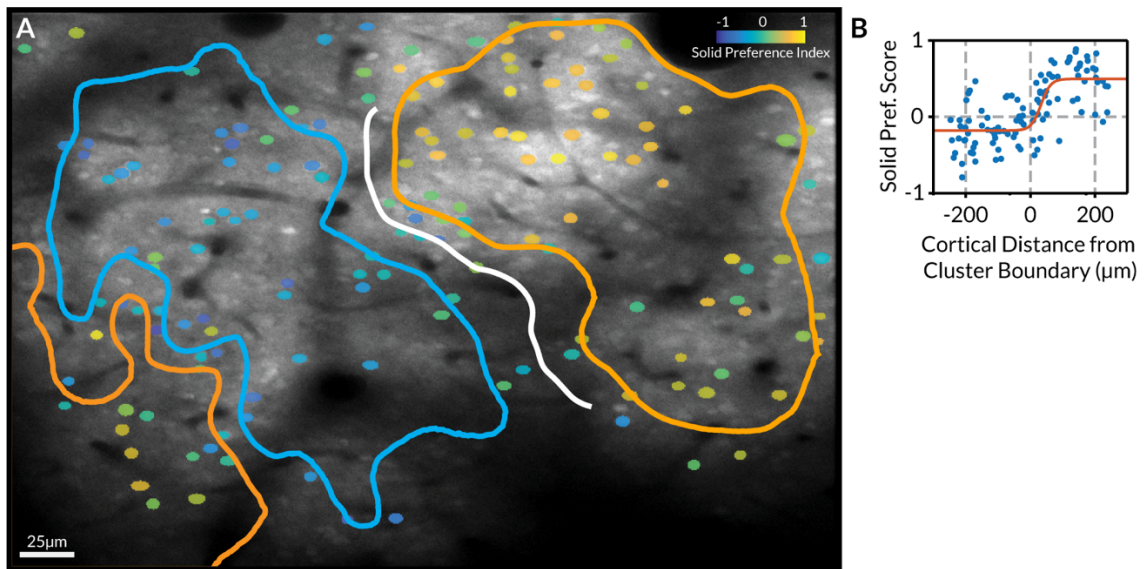


Figure 3.8: Spatial clustering and transition of SP along cluster boundaries.

(A) Imaging region (figure 3.1) overlaid with SP map of neurons (figure 3.6) and selected cluster contours (figure 3.7). Locus of adjacent cluster contours is used as a proxy for the cluster boundary (white) for transition analysis. (B) Transition of SP orthogonal to the cluster boundary (white curve in (A)). Neuron SP values (blue dots) are plotted against the shortest distance from the cluster boundary. Sigmoid function fit to the points is plotted in red.

To ascertain how SP values transition across the boundary between the clusters, a contour that was located in the middle of the two clusters was drawn by calculating the mid-point of the nearest pairs of points along the two cluster contours. This boundary demarcated the boundary between clusters (white line in figure 3.8A). A signed perpendicular distance between every neuron within the region and the cluster boundary was then calculated, and the SP was plotted against this distance (figure 3.8B). For the example region, a sigmoidal function fit to the data shows transitions with a steep slope, so that the SP changed from values ~ 0 to values near ~ 0.6 over approximately $44\mu\text{m}$. To summarize, for the example imaging region, there exist two clusters that have neurons with similar SP values, and those values tend to transition sharply across the clusters.

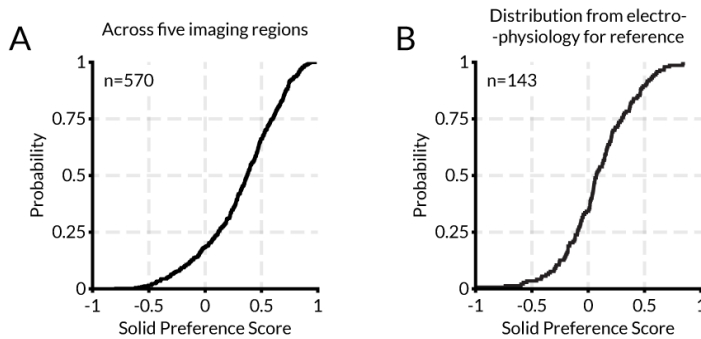


Figure 3.9: Cumulative distribution of SP in two-photon imaging experiments.

(A) Cumulative distribution of SP values across five V4 imaging regions from three monkeys. **(B)** Cumulative distribution of SP from electrophysiology experiments for reference (from figure 2.6A).

A total of five V4 regions were imaged across three monkeys.

The cumulative distribution of SP values across all imaging regions is shown in figure 3.9A, along with the cumulative distribution from the electrophysiology experiments from chapter 2 for comparison (figure 3.9B from figure 2.6A). The pooled

distribution for the two-photon data was significantly greater than 0 ($p < 10^{-80}$; t-test) and had a median of 0.37 and mean of 0.33. 465/570 neurons (81.6%) had an $\text{SP} > 0$, i.e. a greater response to solid shapes on average than planar shapes, and 193/570 neurons (34%) had an $\text{SP} > 0.5$, i.e. a greater than double the response to solid shapes on average compared to the planar shapes. As a comparison, about 66% of neurons had an $\text{SP} > 0$ in the electrophysiology experiment (figure

3.9B). The two distributions were significantly different ($p < 10^{-14}$; Kolmogorov-Smirnov test). A trivial explanation of this difference is that it is a reflection of the varying distributions of SP values across imaging regions. Because two-photon imaging samples only layer 2/3 neurons and the laminar specificity of neurons is indeterminate in single-electrode recordings, it is also possible that the enhanced solid shape preference reflects varying selectivity across cortical laminae.

The SP maps for the other four imaging regions (regions 2-5) are shown in figure 3.9C-F. The cluster contours and the boundaries are overlaid on the SP maps and plots of SP transitions across the cluster boundaries are shown on the right. These data confirm the observations made for the example imaging region: imaging regions consistently show a clear patchy micro-organization of solid shape preference. Imaging regions were randomly chosen across V4, and therefore reveal random subsets of the V4 micro-organization. In addition, imaging regions were relatively small – these data suggest, roughly on the same scale as the solid shape and planar shape clusters. It is therefore not surprising that across all samples, some imaging regions have multiple clusters and clear transitions between them, while other regions are mostly dominated by a single cluster. While this does not impact the conclusion that SP values are clustered in V4, the relative position of solid and planar clusters in an imaging region does limit the analysis of the rate of SP change across cluster boundaries. In addition, the shadows cast by surface blood vessels occlude parts of some of the imaging regions (see figure 3.10C, for example), further limiting the ability to quantify SP changes. Given these circumstances, it is challenging to make a definitive case for either sharp or smooth transitions of SP values across cluster boundaries, and more imaging regions that contain large clusters with clear boundaries are required.

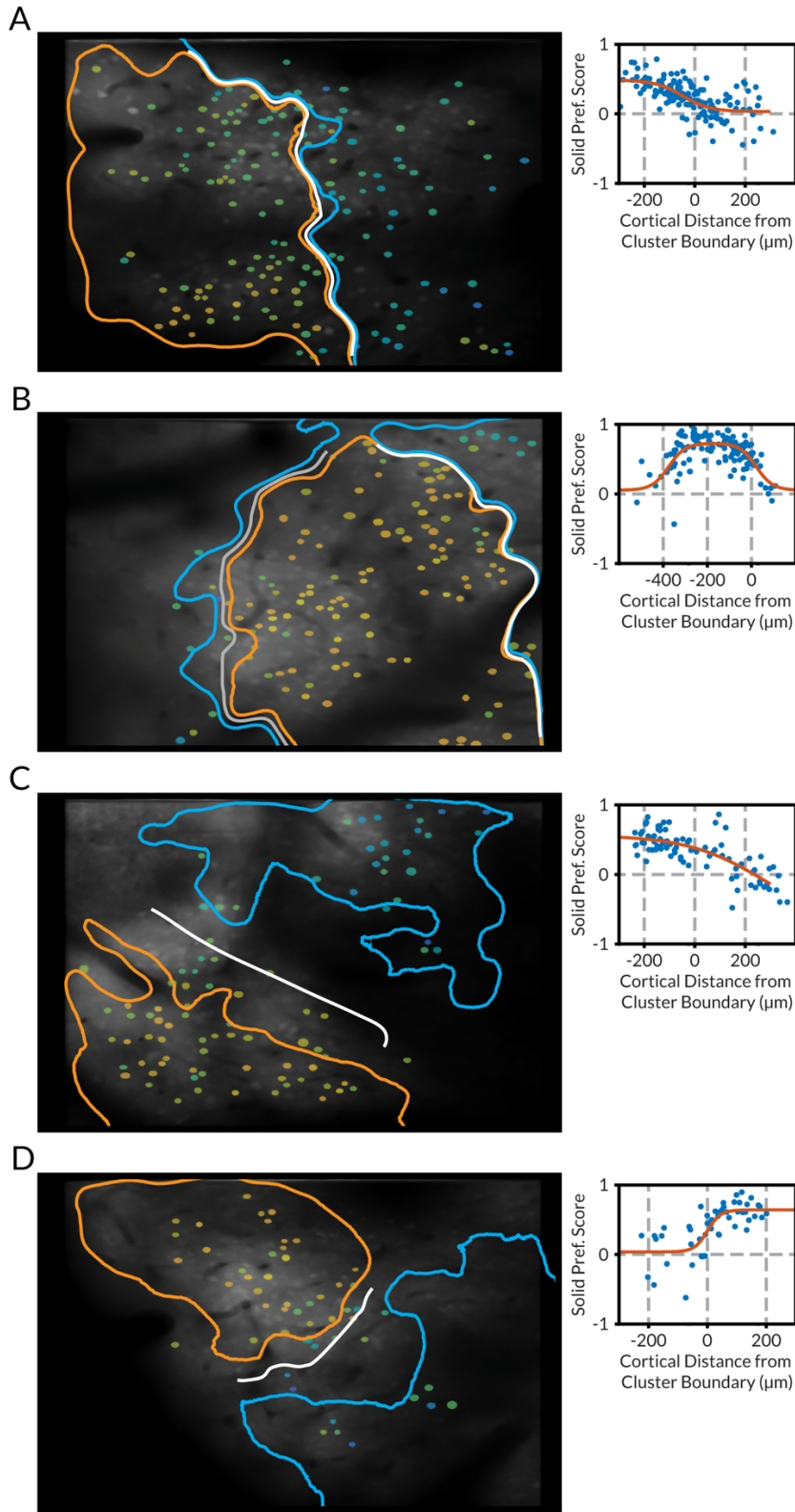


Figure 3.10: Spatial clustering of SP for four more imaging regions.

(A) Anatomical image of an imaging region overlaid with the SP map of the region. The solid shape-prefering cluster (orange contour), planar shape-prefering cluster (cyan contour), and the cluster boundary (white contour) are also overlaid. The transition of SP values as a function of the shortest distance to the cluster boundary is plotted on the right of the image. A sigmoid function (red curve) is fit to the SP values. (B-D) Same as (A) for three more imaging regions. (in (B) a sum-of-sigmoids is fit to the SP vs distance data because there were two planar-prefering clusters flanking a solid-prefering cluster.)

3.2.3 Comparisons of shape preferences within and across clusters

The pixel map-based spatial clustering analysis described above was based solely on the SP of the pixels and the distribution of SP of the neurons inside the cluster. Since SP is based on averages across solid and planar shape stimuli, the clustering was agnostic to the shape-specific response patterns of neurons in the region to the stimulus set. Therefore, the correlations of shape response patterns across neurons within and across clusters can be analyzed to assess the degree to which neurons within and across clusters share shape preferences.

First, the response of the neurons within each cluster was averaged across all neurons for each stimulus yielding an average cluster response vector (see schematic in figure 3.10A and appendix figure 6.5A). This served as a proxy for the response of the cluster. Then the response pattern of each neuron within either cluster was correlated with the averaged response of the solid and planar cluster. The cumulative distributions of these correlations are plotted in figure 3.10 for every imaging region (top row of each set in figure 3.10). In every case, the response correlation of neurons from the solid-preferring cluster with the averaged solid cluster response was considerably higher than the correlation of the same neuronal responses with the planar cluster response (top-left of each set). In fact, this result was the same across all clusters, regardless of solid shape preference ($p < 0.001$ in all cases; Kolmogorov-Smirnov test). In other words, neurons within any cluster were highly correlated with the average response of the cluster they belong to, but not to the average cluster response of the adjacent cluster. This implies that neurons within each cluster do not just share their preference for solid or planar shapes, but additionally share shape preferences. When correlations were calculated for each pair of neurons within each cluster (not the cluster average) the results were similar (appendix figure 6.4).

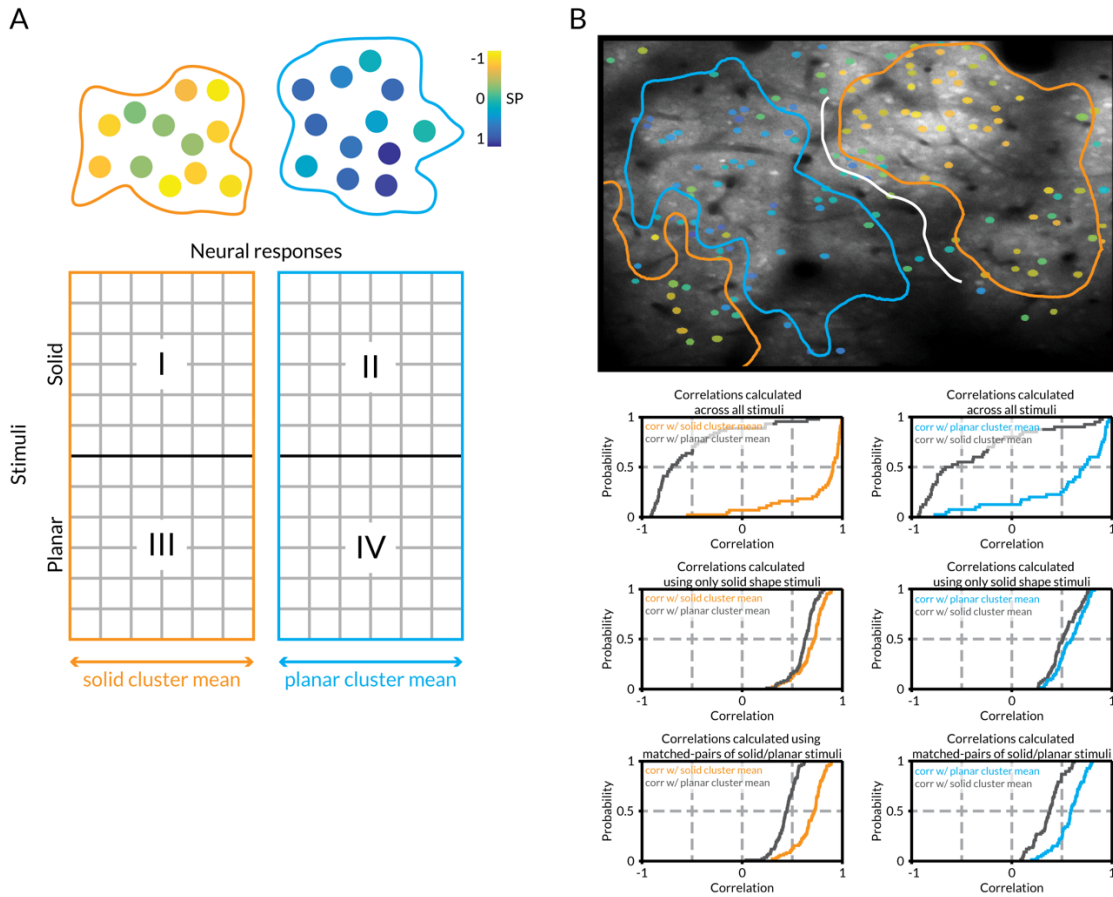


Figure 3.11: Correlational analyses for all five imaging regions.

(A) A schematic depicting the two clusters of neurons preferring solid and planar shapes. The grid depicts a matrix of responses to solid and planar stimuli (rows) for each neuron (columns) in the two clusters. Roman numerals divide the matrix into four quadrants. Correlational analyses in (B-F) are performed on either the full matrix or a subset of quadrants of this matrix. Also see appendix figure 6.5 for additional schematics. (B) For an example imaging region, three types of correlations are plotted in the three rows. Left column (orange and grey curves) represents correlations of cluster average responses with neurons in the solid shape-preferring cluster. Right column (cyan and gray curves) represents correlations of cluster average responses with neurons in the planar shape-preferring cluster. Correlations of neurons with averaged cluster responses across all stimuli (top row; figure 6.5A), only solid shape stimuli (middle row; figure 6.5B), and matched solid-planar shape stimuli (bottom row; figure 6.5C) are arranged below the functional map. (C-F) Same as (B) for four more regions. (... continued on the next page)

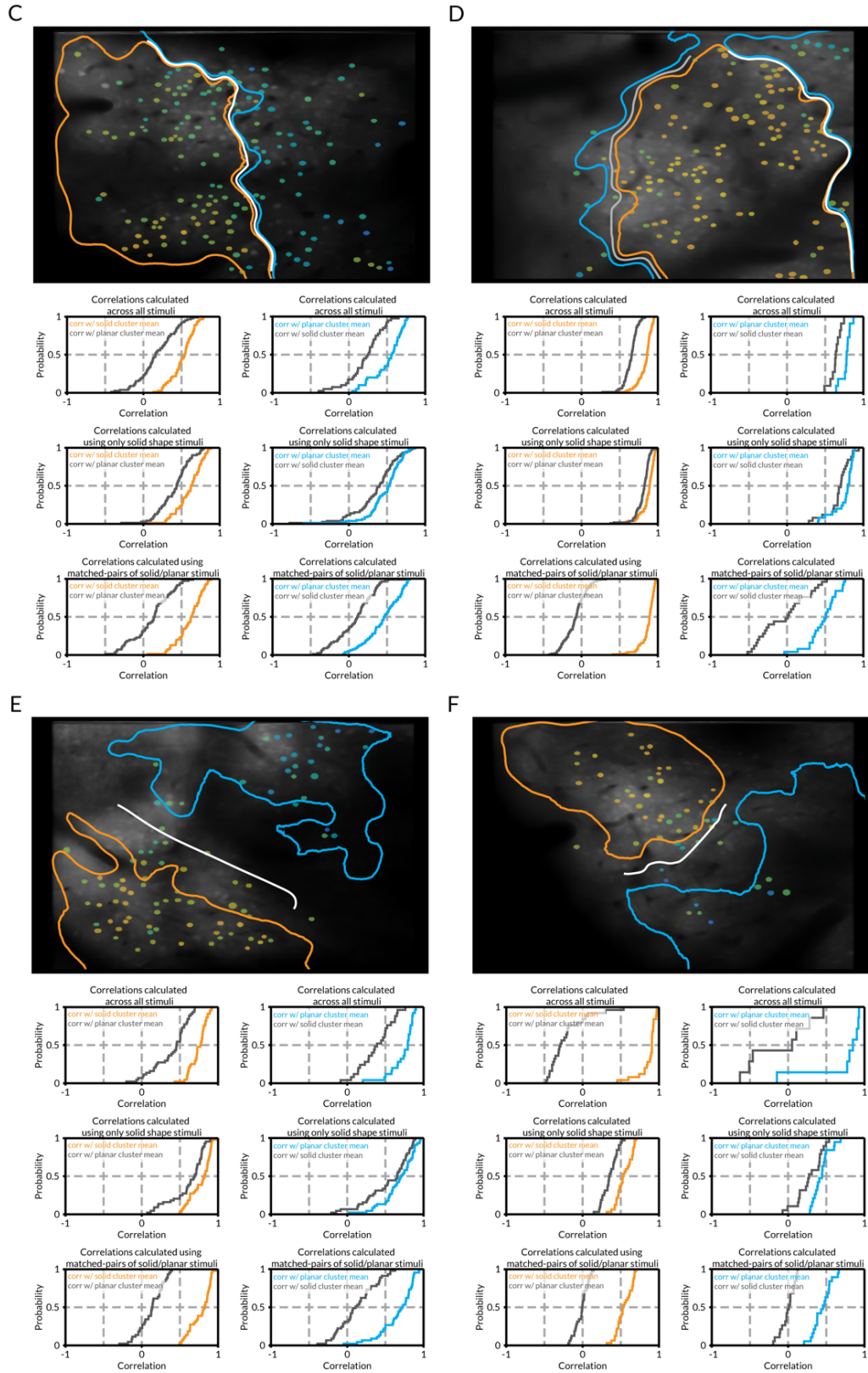


Figure 3.11 (... continued from previous page): Correlational analyses for all five imaging regions.

Next, correlations of only solid shape responses were calculated with the averaged solid shape responses of each cluster, independent of the solid/planar preference of the cluster (middle row of each set in figure 3.10 and schematic in figure 6.5B). This was done to check whether neurons in adjacent solid- and planar-preferring clusters responded similarly to solid shapes. The correlation distributions showed smaller differences indicating that both clusters in every region largely responded in a similar pattern to solid shape stimuli. This difference is especially pronounced in region B where the neurons were largely anti-correlated in the previous test indicating mutually exclusive response patterns but, in this test, became indistinguishable from each other.

Because every solid shape tested had a planar shape counterpart in the stimulus set, a final test was performed to check whether solid and planar shape-preferring clusters responded similarly to the solid and planar versions of the same shape (bottom row of each set in figure 3.10 and schematic in figure 6.5C). In this case, a high correlation with the averaged response of the adjacent cluster would mean that the solid shape responses of the solid shape-preferring cluster follow the same pattern as the planar shape responses of the planar shape-preferring cluster. This test yielded mixed results. In all cases except region B, the planar shape tuning of the planar cluster was different from the solid shape tuning of the solid cluster. As with the SP transitions across boundaries, conclusions from these analyses are limited by the available imaging regions and more data are needed to determine whether there is indeed no correlation between the shape preferences in neighboring clusters, or whether there is a systematic organization of shape preferences in V4 that would be visible at larger scales and/or with more imaging regions. These and other hypotheses are further discussed in section 3.3 and 4.3. In summary, these correlational analyses reveal that neurons within spatially defined clusters have strikingly similar response properties to shapes, in addition to their overall shared preference for solid or planar shapes.

3.3 Discussion

This chapter described the results of two-photon calcium imaging experiments that were designed to probe the micro-architecture of solid shape tuning in V4. Unlike the previous chapter, parametrically defined sets of solid shape stimuli and their planar counterparts were used to access the solid shape preferences of single neurons. The distribution of SP values across five imaging regions recapitulated the distribution of SP values obtained in the previous chapter. Neurons in any given imaging region were found to be organized in clusters of high SP and low SP and these clusters were found to be juxtaposed. In addition, the shape preferences of neurons within each cluster were found to be highly correlated. These results are the first demonstration of an orderly micro-organization of stimulus properties in V4. This discovery further reinforces the results from the previous chapter that V4 neurons are tuned for solid shape fragments; the observation that neurons are not only tuned to solid shape fragments, but that solid shape preference is mapped in V4, is strong evidence for the hypothesis that solid shape is a *critical feature* in V4. This micro-organization has implications for coding transformations within V4, raises questions about the mechanisms by which such a mapping might be achieved, and sheds light on how downstream areas might be tuned for solid shape. These, and other implications, are discussed below.

In the imaging regions illustrated in this chapter, the overall size of the clusters in V4 is roughly 130-180 μ m in diameter. This estimate comes from the clusters in the regions in figure 3.8 and figure 3.10B where a central solid (or planar) cluster is flanked by two planar (or solid) clusters. This estimate is preliminary because at the spatial resolution at which neurons were clearly visible and yielded high signal to noise ratios, imaging region were sized such that they typically contained only one or parts of two clusters. Nevertheless, the fact that juxtaposed clusters were observed in most regions is evidence for the hypothesis that the clusters are typically at most the

size of the imaging region, if not smaller. Further experiments at larger spatial scales are required for a better estimate.

While preliminary, these data suggest that solid shape preference sharply changes across cluster boundaries. This would be consistent with an organization that consists of segregated modules for solid and planar shape-selective neurons. This conclusion currently is based on the sharp transitions in three out of five imaging regions and should be validated with more imaging samples and/or imaging larger regions that are more likely to contain several cluster boundaries.

Furthermore, the solid shape preferences of neurons within each cluster were tightly correlated, at least for the limited stimulus set used here. Based on the precise tuning curve measurements obtained in chapter 2, the responses of neurons observed for the simpler stimulus set can be explained by tuning for 3D object-centered position, curvature, tubular width, junction angle, etc. Therefore, the result that solid shape preferences are clustered in V4, raises the interesting hypothesis that the underlying tuning functions are also clustered. To test for shape tuning of single neurons and the mapping of shape tuning in V4, correlational analyses are not sufficient, and a larger stimulus set would be required. This larger set may take the form of an adaptive shape morphing algorithm, like the one used in the previous chapter. Instead of sampling the tuning manifold of a single neuron, this algorithm would need to optimize for the mean population responses, differences of responses of clusters of neurons, and/or the response density of the population. In addition, the larger amount of sampling required for these experiments would only be possible in a different preparation that allows longer optical access to each imaging region, such as chronic two-photon imaging.

In discussing the correlation of shape preferences between neurons in a cluster, it should be noted that the manual identification of cell bodies in each imaging region opens the possibility of

contamination of neural signals by signals originating in the neuropil. Since the neuropil signal appears to share tuning properties with the neurons that are embedded within it, this contamination has the potential to artificially increase correlations between neurons. However, the observation of sharp transitions in SP across cluster boundaries argues against strong neuropil contamination. In addition, previous analyses of the relationship between neural responses and the neuropil around them has been quantified before (Nauhaus et al., 2012), and could be analyzed here as well. In addition, existing algorithms for removing neuropil contamination could be applied to further reduce spurious correlations.

3.3.1 Relationship to known maps in V4

Cortical maps in V4 have previously been studied using optical imaging techniques paired with electrophysiological recordings. These studies demonstrated maps of color/hue (Tanigawa et al., 2010), orientation (Ghose and Ts'o, 1997), and disparity-defined edges (Fang et al., 2018). All maps typically revealed patchy organization of the stimulus properties and haphazard mutual organization among the maps. Given the results of this chapter, the mutual organization of solid shape preference and other known cortical maps in V4 would be the next frontier of investigations. The previous maps of orientation, color, and disparity were obtained with low resolution techniques. It will be interesting to redraw these maps with single cell resolution and investigate the joint tuning in single neurons and clusters for solid shape preference and other stimulus properties.

In this study, only the map of solid shape preference was obtained. In order to align this map with those for orientation, color, and disparity, further experiments would be required which either factorize the relevant solid shapes with these properties to produce solid shape stimuli that vary in color, depth, etc. or acquire the maps of solid shape preference and other properties in separate experiments.

The results in this chapter address the cortical map in layer 2/3 only. Due to the density of tissue staining and the opacity of the brain, two-photon imaging cannot reach deeper cortical layers and therefore cannot make inferences about columnar organization. Electrophysiological studies that evaluated the change in neural tuning to color perpendicular to the cortical surface reported a vague columnar organization in V4 (Kotake et al., 2009). Testing whether a columnar organization of solid shape tuning exists in addition to the observed horizontal organization will be an exciting prospect for future experiments using either laminar probes or three-photon imaging. This would also allow a quantification of how the neural tuning to solid shape fragments evolves between layers.

3.3.2 Implications for V4 inputs and outputs

As mentioned in the introduction, a slight rotational difference in the mosaics for on- and off-RGC may be sufficient to cause a bias in LGN inputs to V1 that in turn seeds the map of orientation tuning. By applying the idea of using spatial pooling of inputs to explain feature maps in an area, the solid shape map found here in V4 may be used to inform circuit models of the ventral stream. So far, models of the pathway have had to rely on explaining how combinations of tuning properties in one area could lead to the tuning properties in the next area. Adding the spatial layout of feature maps in an area (or in the case of V1 and V4, the feature maps in the input and output area) imposes important additional constraints on these models. These spatial constraints will generate more specific circuit models, that can make testable predictions e.g. about the spatial footprint of V1 projections to V4. Similarly, does the existence of a solid shape map in V4 have implications for coding transformations in downstream areas? If indeed a map of solid shape tuning exists, pooling from adjacent clusters of solid shape preference in V4 would be an easy mechanism to create neural tuning to complex combinations of solid shapes as is observed in IT cortex (Brincat and Connor, 2004; Hung et al., 2012; Yamane et al., 2008).

In conclusion, the micro-organization of solid shape in V4 raises many questions about the implementation of coding algorithms in the ventral visual stream. In light of the results in this chapter, the local and global circuits that process the visual world reinterpret the incoming signals into 3D reality in the early stages using mechanisms yet to be discovered (see general discussion in chapter 4 for more details).

3.4 Methods

3.4.1 Experimental Model and Subject Details

Three juvenile male rhesus macaques (*Macaca mulatta*) aging between 15-21 months and weighing between 3.0 and 4.0 kg were used for anesthetized two-photon imaging experiments. They were singly housed before the experiments and were not involved in any prior procedures. All procedures were approved by the Johns Hopkins Animal Care and Use Committee and conformed to US National Institutes of Health and US Department of Agriculture guidelines.

3.4.2 Method Details

Animal Preparation and Surgery

The monkey was anesthetized with ketamine (10mg/kg IM) and pre-treated with atropine (0.04mg/kg, IM) to stabilize heart rate. A tracheostomy was performed and an endotracheal tube was inserted to deliver oxygen and isoflurane anesthesia, starting at 3% and lowered to ~0.5% during imaging experiments. An intravenous line was inserted for delivering IV drugs and a Lactated Ringers Solution (LRS) containing 5% dextrose. The animal was placed in a stereotaxic apparatus and the head was initially secured using ear bars, eye bars, and a palate bar. Later, a titanium head post was implanted using cranial screws and dental cement to secure the head to the stereotaxic apparatus for the rest of the experiment. At this point, the ear bars and eye bar were

removed. The metal head post and the palate bar setup were most efficient in controlling breathing related artifacts during imaging experiments. Cranial metal screws were implanted over the frontal lobe to record the EEG. Throughout the experiment, anesthesia was maintained with sufentanil citrate (4-20 μ g/mg/h, IV), supplemented with low levels of isoflurane. The animal was paralyzed with pancuronium bromide (0.15mg/kg/hr, IV) and artificially ventilated with a small animal respirator (Ugo Basile). Dexamethasone (0.1mg/kg, IM) and cefazolin (25mg/kg, IV) were administered to prevent infections and swelling. EEG, EKG, SpO₂, EtCO₂, respiration and heart rate, and body temperature were monitored continuously to maintain the appropriate depth of anesthesia and monitor animal health throughout the experiment.

The skull over dorsal V4 was thinned to reduce the distance between the imaging objective and the cortical surface. After estimating the location of the lunate sulcus and the superior temporal sulcus based on stereotaxic markers, a disc-shaped metallic well was cemented over V4 and parts of V1. Subsequently, diagnostic craniotomies were performed to locate the two sulci and confirm the location of V4. Small craniotomies and durotomies (~9 sq. mm.) were then made to expose V4. The cortical surface was visualized with epifluorescence imaging. A dye solution of 2mM Oregon Green BAPTA 2-AM, 10% DMSO, 2% pluronic, and 25% sulforhodamine 101 (ThermoFisher Scientific) in ACSF was loaded into a glass pipette, with was lowered into cortex at an ~ 45° angle using a micro-manipulator (Sutter Instruments) under microscopic guidance. Several injections of the dye solution were made into a single durotomy using a Picospritzer pressure injection system (Parker Hannifin). After a ~1hr waiting period, the craniotomy was covered with 1.5% Type III Agarose (Millipore Sigma) and a glass coverslip (Warner Instruments) was inserted under the craniotomy to reduce vertical motion artifacts due to breathing. Horizontal motion artifacts were removed post-hoc. Most imaging data was collected at cortical depths of ~100-200 μ m. After the imaging experiment ceased to yield results due to photobleaching of the cortical region or if the dye loading failed to produce sufficient well-

labeled neurons, a new craniotomy was performed, and the dye loading was repeated in the new location. Each hemisphere typically yielded 4-6 craniotomies after which the imaging well was re-located to the other hemisphere.

Before the imaging experiment, both eyes were covered with contact lenses to protect them from drying. Refraction of the eyes was determined for the stimulus display at 60cm from the eyes based on electrophysiological recording of a patch of V1 cortex. In this experiment, neural responses to oriented gratings at several spatial frequencies were recorded with ophthalmic lenses of different strengths placed in front of each eye. Lenses which optimized responses to the highest spatial frequency gratings were then chosen for each eye.

Two-photon microscope

Two-photon microscopy was performed using a NeuroLabware microscope coupled to a Chameleon 2 Coherent Ti-Sa laser. A 16X water-immersion objective (Nikon; 0.8 NA lens and 3mm working distance) was mounted on a movable stage with one rotational and three translational degrees of freedom for easy placement perpendicular to the imaging region. Imaging was performed at 920 nm and the emission was collected at 15.5Hz using green and red photomultiplier tubes.

Visual Stimulus Generation

Unlike the single-electrode physiology experiment described in chapter 1, the two-photon imaging experiment was time-limited to a few hours per imaging region and, therefore, not amenable to precise mapping of solid shape tuning of each neuron in the imaging region. A constrained set of solid and planar shape stimuli constructed was used to map the relevant solid shape properties, as illustrated in figure 3.3 and 3.4. All stimuli were generated with OpenGL rendering of solid shapes with a single light source placed above the virtual camera. The 3D

model for the solid shape stimuli was generated using a precise mathematical shape generation algorithm with an ability to change 3D medial axial and surface properties during the experiment. Flat, planar stimuli were generated by disabling OpenGL lighting.

Visual Stimulus Presentation

Stimuli were presented using the Psychophysics Toolbox (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997) extension for MATLAB (Mathworks Inc.). The toolbox was also used to synchronize visual stimulus presentation with two-photon image acquisition. The stimuli were presented on a 120Hz 24" LCD display (Viewsonic) gamma-corrected using a Photo Research spectroradiometer. A set of receptive field (RF) localizing stimuli (gratings, bars, 2D sprites) were presented and the average receptive field of neurons in the imaging region was determined using pixel-based online analysis (figure 3.2). The stimulus display was roughly centered on the RF location.

In one animal, visual stimulus set 1 (C-shaped and L-shaped, see figure 3.3) was flashed within the RF. In the other two animals, visual stimulus set 2 (long, bent tubes, see figure 3.4) was drifted across the full field of view. Stimuli in set 2 were drifted across the screen at a uniformly sampled set of directions such that every pixel encounters precisely 2 cycles of the visual stimulus. Additionally, the stimulus was displaced in a direction orthogonal to the direction of drift in 11 increments such that each fragment of the solid/planar shape was drifted across the RF of each neuron.

For both stimulus sets, a smaller set of coarsely sampled stimuli were chosen for initial experiments in each imaging region. For every experiment, the relevant stimulus set was randomly shuffled, and each stimulus was presented for 1-2s followed by a blank gray screen for 1-3s (see time courses in figure 3.5). Five blank trials (gray screen) were also randomly presented

during the experiment. The long stimulus and inter-stimulus durations was matched to the slow time course of OGB responses and acquisition rate. Each stimulus was repeated 5-10 times.

Image pre-processing and segmentation

An anatomical reference image was created for each imaging region by averaging uniformly sampled frames across the experiment duration. For longer experiments, several reference images were created. For in-plane movement correction, each frame was correlated with the reference image and a horizontal and vertical shift was found. These shifts were applied to the frame to yield a stable recording. This correction was applied to each frame before further analyses. A binary cell segmentation mask was created for the imaging region by manually encircling cells in the reference image. Glia cells were accounted for by selective labeling with SR101. Thus, cells present in both red (575-650nm emission filter) and green (510-560nm emission filter) channels were rejected. The response of each neuron in each frame was computed by the dot product of the movement-corrected frame with the neuron's binary mask. Each imaging region yielded ~25-200 neurons.

The stimulus onset and offset times were tracked in reference to two-photon scanning. Since the laser scanning started from the top left of the frame and proceeded across the image row-by-row, neurons were scanned in a sequence from top to bottom. The onset of the stimulus could occur when the laser is in the middle of scanning a frame. Therefore, it is not sufficient to synchronize the frame number with the stimulus onset. Each neuron's response time-course was therefore corrected based on its location in the image and the exact onset of the stimulus, i.e. if the stimulus appears in the middle of the scanning of a frame, that frame was only used as the response of neurons that were scanned after stimulus onset. The fluorescence time-course ($\Delta F^n(t)$) of each neuron for each stimulus trial was calculated as:

$$\Delta F^n(t) = \frac{f^n(t) - F_o^n}{F_o^n}$$

where $t = 0$ represents the stimulus onset, $f^n(t)$ is the measured fluorescence response of the n^{th} neuron, and F_o^n is the average response of the neuron in the baseline period ($t > -1200ms$ and $t < -100ms$). The stimulus response of each neuron was calculated as the average $\Delta F^n(t)$ for $t > 100ms$ and $t < t_{off} + 600ms$ (where t_{off} is the time of stimulus offset).

3.4.3 Quantification and Statistical Analysis

For imaging regions with stimulus set 1, only cells that passed a t-test ($p < 0.05$) between the best and worst stimulus condition were included in further analyses. Because this statistic was only used for rejecting neurons that were unresponsive to any stimuli in this set and no additional per-neuron or population analyses were performed based on this rejection, this statistic was deemed appropriate. For imaging regions with stimulus set 2, since each stimulus drifted across the screen in each trial (time) and was displaced in the orthogonal direction across trials (space), this created a 2D matrix of responses (similar to the event map studied extensively in the somatosensory cortex). To effectively capture the significance of neural response across time and spatial shifts, a two-sample, two-dimensional Kolmogorov-Smirnov test (Justel et al., 1997) was calculated across time and stimulus displacement between the best and worst stimulus condition. Only cells that passed the 2D KS-test ($p < 0.05$) were included in further analyses.

Solid Shape Preference Index Map

As with the analysis in the previous chapter, the solid preference index for each neuron in an imaging region was calculated as:

$$SP = \frac{R_{solid} - R_{planar}}{\max(R_{solid}, R_{planar})}$$

where R_{solid} is the average neural response across all solid stimuli and R_{planar} is the average neural response across all planar stimuli. This score varies between -1 (prefers only planar stimuli) to 0 (no preference) to +1 (prefers only solid stimuli).

Spatial Clustering Analysis

To group neurons in an imaging region into meaningful spatial clusters, a pixel map of SP was created (see figure 3.5) i.e. each pixel was assigned a SP score using the same formula as above. The map was then smoothed with a broad 2D Gaussian kernel using the in-built MATLAB function *imgaussfilt* with a standard deviation of 20, normalized, and a contour map was generated. The 2D contours from this contour map provided cluster candidates. All contours went through a 3-stage selection/rejection criterion. *Stage 1*: Contours that contained fewer than four neurons or were smaller than 10% of the size of the imaging region in area, or were too complex (ratio of the area of the contour to the area of the convex hull of the contour was less than 0.95) were rejected first. *Stage 2*: For remaining contours that contained the same neurons, the most complex contours were rejected. *Stage 3*: For the remaining contours, concentric contours were identified and the largest contour with the least SP kurtosis (fourth moment of the distribution of SP of neurons within a contour) was chosen. This ensured a homogeneous distribution of SP values within the selected clusters. In some cases, after the contours were selected, an additional contour simplification step was done where the contour was replaced with its convex hull. This was not done when the simplified clusters intersected each other.

Boundary Transition Analysis

The boundary between adjacent clusters was defined as the set of points that are equidistant from the two cluster contours. To calculate this, the pair of closest points on the adjacent cluster contours was identified and the mid-point of their connecting line segment was calculated. Both contours were then traversed in both directions and mid-points were calculated for all closest point pairs. Finally, the cluster boundary was defined as the smooth b-spline curve created with these mid-points as control points.

4 Discussion

The results in this thesis address fundamental questions about coding transformations in the ventral stream by focusing on processing of solid shape information in intermediate area V4. In chapter 2, a large fraction of V4 neurons recorded extracellularly was found to respond preferentially for solid shapes versus planar shapes. Response-weighted averaged (RWA) matrices suggest that these neurons have tuning peaks with dimensions for 3D orientation, curvature, radius of tubular structures, junctions between two solid limbs, 3D object-relative position. The responses to solid shapes were consistent across lighting cues like shading and specularity, and optical cues like reflectivity and refraction. Also, neurons retained their preference for solid shapes when rendered as random dot stereograms (RDS) without any monocular contrast or texture cues. In chapter 3, the distribution of solid shape selectivity was recapitulated in two-photon calcium imaging experiments. In addition, neurons were clustered according to their preference for solid or planar shapes, with sharp boundaries between the clusters. Finally, within each cluster, the shape preferences of neurons were highly correlated. The major contribution of this thesis is the discovery of solid shape representation at an early visual processing stage that was thought to only process image information. These results necessitate a rethinking of the hierarchy of coding transformations along the ventral visual pathway. This finding may also provide a potential resolution to a longstanding argument about coding of contour, spatial frequency, texture, and disparity information in V4. These previous results, when viewed from the perspective of solid shape coding, resolve into cues that are informative about the solid shape of an object. Indeed, the results in this thesis argue that the ventral visual pathway transforms the input image into an intermediate (or latent) code which is decipherable in terms of explicit solid shape representation to efficiently extract the identity of the object in the image. In the following section, the concept of latent representations is analyzed in view of the current results and deep network models of object vision.

4.1 Image \Rightarrow Solid Shape \Rightarrow Identity

The *identity* of an object is not limited to a set of category labels. The words ‘a pen precariously placed at the edge of a table’ conjure up a mental picture that is not fully describable by the very sentence that conjured it. ‘Pen’ and ‘table’ and maybe even ‘precarious’ may be gleaned by a computer using a photograph of the scene, but inferences about the scene that the brain produces – how one would shake the table to make the pen fall, how one would catch the pen as it is falling and whether or not one would be able to get there before it hits the ground – would require millions of training images and trials for a computer to generate. The shapes of the objects in the scene are a fundamental aspect of the way these inferences are made with visual information.

Inherent in an object’s visual shape are properties like balance, animacy, weight, affordance, and capacity. The representation of these properties requires a detailed representation of object shape. A novel object with a handle informs us how the object is potentially lifted with no instruction. If the same object is asymmetrically shaped, we can guess that it will be unevenly weighted. Shapes are central to visually guided interactions with objects in the world.

4.1.1 Deep network models of object vision

It is a matter of debate as to how these inferences are performed in the brain. The current best class of mechanistic models of object vision – deep learning models – are trained on classification of objects into categories given an image (Cadena et al., 2017; Cadieu et al., 2014; Cichy et al., 2016; Khaligh-Razavi and Kriegeskorte, 2014; Schrimpf et al., 2018). These models stack convolutional and normalization layers to achieve human-level performance at classification (see figure 4.1). They assume that hidden transformations convert 2D information directly into object identity (He et al., 2016; Huang et al., 2017; Krizhevsky et al., 2012; Simonyan and Zisserman,

2014). Importantly, it is believed that the inner representations (or latent space in figure 4.1) of the deep network are not describable as the representation of meaningful parts of the image or the object within the image i.e. there are no reasonable geometric parameters that can be used to model the responses of single units within these inner representations. Similarly, in the ventral visual pathway, many studies have considered the description of the critical features in intermediate visual cortex to be a futile endeavor (Abbasi-Asl et al., 2018; Bashivan et al., 2019). These studies instead suggest that the trajectory of representations between early visual cortex (V1) and late visual cortex (IT) is indescribable, except in a strictly mathematical sense as a set of complex image transformations. The results presented in this thesis take an opposing view, in that they attempt to describe that intermediate representational space in the brain. The findings demonstrate that V4 produces explicit signals for the solid shape geometry regardless of the 3D cue.

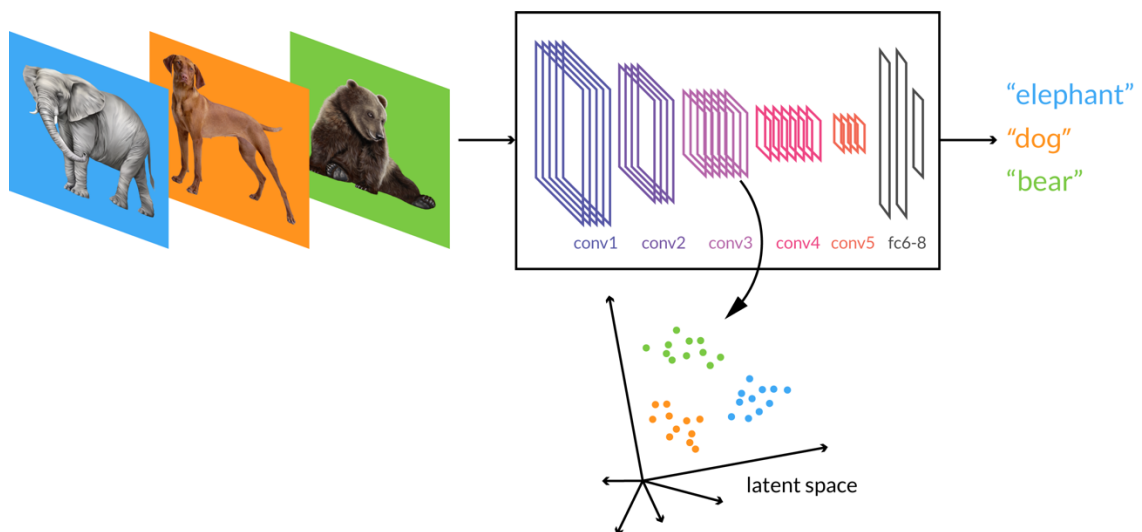


Figure 4.1: Neural networks and latent spaces.

Deep convolutional networks (DCN) are neural networks that perform 2D convolutions on the input image with learned kernels to sequentially transform the image into a latent space to aid clustering and classification of the image. In the black box, the architecture of a popular DCN called AlexNet is shown. The convolutional layers of this network been successful as models of neurons in the ventral visual stream.

The result that the 3D reality of an object is extracted before its identity has many implications. Firstly, shape representation is a more stable code than retinotopic representation because it is robust to small changes in viewing conditions. In other words, if, via coding transformations in the early- and mid-level areas, object shape is explicitly identified first, the extraction of object identity in later stages will potentially be less vulnerable to changes in viewing conditions, global shifts in illumination, occlusion, etc. The space of planar images is highly degenerate, in the sense that the same 2D silhouette could be produced by many solid shapes. A more abstracted representation for solid shape fragments would produce a sparser and more explicit representation, which would aid in transfer and decoding. Furthermore, because edge information is preserved in V4 in neurons that prefer planar shapes over solid shapes, this may serve to robustly represent information about texture and other planar information like text on solid shapes (Ratan Murty and Arun, 2017). Indeed, the result that solid and planar shape preferring neurons are clustered next to each other, perhaps forming two channels of processing, may contribute to the joint robustness of solid shape discrimination and edge and texture discrimination in primates. The lack of similar processing channels, in turn leading to a lack of robustness against changes in viewing conditions, may be the reason for the susceptibility of computer vision algorithms to attacks that are imperceptible to humans and other primates.

AlexNet (Krizhevsky et al., 2012) is a popular deep convolutional network (DCN) that has been successful in modelling neural responses in V4 and IT using natural and synthetic images (see figure 4.1). When trained on neural data, a similar deep network model produced units that behaved quantitatively like V4 units (Cadieu et al., 2014; Yamins et al., 2014). These units have since been used to generate testable hypotheses about V4 neural responses by creating synthetic stimuli that drive the neural population in predicted directions (Bashivan et al., 2019). In

experiments not described in this thesis, AlexNet was used as a comparison against the results obtained in V4. The activations of individual unique kernels within AlexNet were correlated with V4 neuron response to the same images that were used to access solid shape tuning in the electrophysiology experiment. This analysis showed that single convolutional kernels in AlexNet responded very similarly to single neurons in V4. An adaptive sampling algorithm, similar to the one run for V4 neurons in the electrophysiology experiment, was also run on the model units, and yielded similar results. This comparison demonstrates that the brain and AlexNet both choose a very similar trajectory of coding transformations, through a latent space describable in terms of solid shape representations of 3D reality, to infer object identity from an image. This surprising convergence lays the groundwork for future comparisons of V4 representations with deep network representations.

4.2 Future directions

4.2.1 Population coding and dynamics of solid shape representation

This thesis described a novel representational space in V4 on the basis of the tuning properties of individual neurons. It also addressed some aspects of the local micro-architecture of V4 neurons. But how responses to contour or surface elements are integrated by local circuits remains unknown. Theories about integration extrapolate from psychophysical experiments and Gestalt principles (Koffka, 2013; Wertheimer, 1938) to suggest population coding theories based on natural shape statistics (Roelfsema, 2006; Roelfsema and de Lange, 2016). Briefly, during development and perceptual learning, neural networks learn co-occurrence statistics in natural images and therefore form a rapid and robust mechanism to integrate these elements (Geisler et al., 2001). This is called base grouping. As a consequence, an auxiliary mechanism is required that allows for the integration of statistically uncommon elements (Jolicoeur et al., 1986; Roelfsema and Houtkamp, 2011). This is called incremental grouping and is more flexible but

slower than base grouping. These perceptual mechanisms have been extended for solid shape geometries by (Dorai and Jain, 1997; Nelson and Selinger, 1998; Selinger and Nelson, 1999) by suggesting that the grouping of solid shape objects requires a hierarchy of transformations for solid shape reasoning and manipulation, reminiscent of the levels of coding transformations along the ventral visual pathway.

It is conceivable that the neural correlates of the perceptual phenomena described above would be visible in the population code of ventral visual stream areas. In other words, neurons that encode shape properties that co-occur more often may be strongly interconnected to amplify their mutual responses to make the visual percept more robust. Because of its position in the ventral visual pathway, receptive field sizes, and tuning properties, V4 is the ideal candidate for studying these mechanisms. Using the orientation and curvature tuning and responses of single neurons in V4, researchers were able to re-construct the entire presented stimulus (Pasupathy and Connor, 2002). This study not only showed that contour orientation and curvature are encoded in V4 neurons, but also that the entire shape can be decoded from V4 population activity. Therefore, it is reasonable to assume that contour integration could occur in V4. Furthermore, V4 response dynamics hint at a non-linear summation of multi-element responses in the late response period which could be the signature of base grouping (Yau et al., 2013). The population coding and multi-element responses have not been investigated in terms of V4's solid shape representation. These results have provided critical insight into the visual representation characteristics at an intermediate visual processing stage, but how V4 local circuits integrate this information remains an open question. Also, the population coding and multi-element response dynamics of V4 neurons have not yet been studied in the context of solid shape responses but may underlie similar integration mechanisms as for planar contours. Testing these theories would require simultaneous measurements from multiple neurons to compare the response dynamics of neurons that have similar and dissimilar tuning preferences. This may be possible with fast-scanning two-photon

imaging of a large region of V4 such that many distinct types of tuning properties can be compared, and/or with multi-channel silicon probes with hundreds of recording contacts.

4.2.2 Local and global circuits for solid shape processing

Orientation tuning is one of the critical features in V1. The leading hypothesis for the formation of orientation-tuned cells in V1 is that on- and off-center thalamic inputs to a single cortical column in V1 are aligned in such a way that the population receptive field of the input resembles the orientation tuning of the column. Evidence for this hypothesis comes from paired recording of neurons in LGN and V1 in cat (Ferster et al., 1996; Reid and Alonso, 1995), ferret (Chapman et al., 1991), and monkey (Jin et al., 2011). While areas beyond V1 also contain orientation tuned cells, all evidence points to orientation tuning being an *emergent* property in V1. Perhaps as a signature of this, V1 contains strikingly orderly orientation maps.

Chapter 2 describes solid shape tuning in single neurons in V4 and chapter 3 describes the organization of solid shape preference in V4. Particularly the observation of orderly maps for solid shape preference in V4 suggests that the theories developed for input integration in V1 might be extendable to explain V4 processing. Just as pooling LGN inputs produces orientation columns in V1, selectivity for contour combinations may be produced by pooling inputs from orientation-tuned columns in V1. Furthermore, simultaneously pooling inputs from disparity-tuned neurons in V2 and/or orientation-tuned binocular neurons in V1 could induce tuning for solid shape-in-depth and 3D orientation. Solid shape responses in V4 are consistent across other 3D cues like reflection and refraction, in addition to shading and disparity. What additional stimulus properties represented in early areas contribute to this are yet unknown. It is possible that the representation of reflective/refractive solid shapes underlies neurons tuned for surface textures which have long been known to be represented in V1 and V2 (Knierim and van Essen, 1992; Ziemba et al., 2016) and pooling across texture-tuned neurons and 3D orientation-tuned

neurons could give rise to the observed cue invariance. Together, these pooling mechanisms could possibly produce representation for solid shape tuning robust to changes in 3D cues in V4. Evidence from paired recordings from V1 and V4 reveal that simple pooling mechanisms can explain the size and location of receptive fields of V4 neurons based on the V1 inputs it receives (Motter, 2009).

An alternate hypothesis is that solid shape tuning is not emergent in V4 but is instead inherited from inputs from V1 and V2. This claim is easily tested by running similar battery of experiments as described in chapter 2 except for neurons in V1 and V2. However, extending the finding that V1 feature maps exist for tuning properties that emerge in V1, the observation that solid shape preference forms a map in the output layer 2/3 of V4 strongly suggests that solid shape tuning is computed in V4.

In the same vein, another open question is that of columnar organization. The electrophysiological recordings in chapter 2 were not targeted to particular cortical layers. The two-photon imaging regions in chapter 3 were restricted to layer 2/3. Therefore, it remains to be tested how solid shape preference and tuning evolves across V4 layers, and whether a columnar organization exists. Experiments that studied the dynamics of shape processing in V4 revealed that tuning for curvature components does not appear immediately, but rather evolves from simpler elements gradually over ~ 50ms after stimulus onset (Yau et al., 2013). This temporal evolution may be a signature of laminar processing where the early response represents the linear summation of the feedforward inputs which are processed along a column to gradually give rise to a non-linear response (Brincat and Connor, 2006; Chance et al., 1999; Yau et al., 2013). Further analysis of the dynamics of solid shape tuning is required to reveal the laminar computations that give rise to the observed solid shape tuning. Future experiments using silicon

multi-site probes inserted perpendicular to the pial surface or 3-photon imaging would be able to access the evolution of solid shape responses along a cortical column.

Alternatively, the observed linear/non-linear dynamics may be a signature of recurrent computations across cortical columns. The leading hypothesis for robust object identification in difficult viewing conditions, like occlusion, is that the neurons that code for the visible and discernable parts of the object are recurrently connected with the invisible parts of the object (Kosai et al., 2014; Roelfsema, 2006). In that case, the solid shape response dynamics could be modeled as an evolving non-linear summation of solid and planar inter-columnar recurrent inputs. If further research reveals that shape tuning is mapped in V4 along with solid shape preference, and planar and solid shape preferring regions for similar shape fragments are juxtaposed in cortex, then it is conceivable that evolving response dynamics are a reflection of recurrent computations that serve to support the robustness of the solid shape representation.

These theories also relate back to potential surface integration mechanisms discussed in the previous section. It is possible that the micro-organization of solid shape tuning across V4 is on the basis of co-occurrence statistics of surface fragments in natural environments. If future experiments do indeed find that to be the case, then that would be critical evidence in favor of (a) natural shape statistics providing strict constraints for cortical organization, (b) recurrent computations amplifying the robustness of solid shape responses by integrating across neurons with tuning for related shapes, (c) cortical organization being responsible for the slow integration of statistically irregular object parts, and related hypotheses.

To test these hypotheses, a region of V4 large enough to contain multiple solid shape preference zones would need to be imaged with an objective with a large field-of-view and fast scanning.

Experiments that pioneered these technological advancements were able to image a 3D volume of cortex up to 5mm wide and 1mm deep with single neuron precision in mice (Ji et al., 2016; Sofroniew et al., 2016; Stirman et al., 2016). These techniques could prove to be central to resolving the critical question of how solid shape preference and tuning are mutually organized in V4.

It is known that a major recipient of feedforward projections from V4 neurons is IT cortex, specifically PIT/TEO (Desimone et al., 1980; Distler et al., 1993; Nakamura et al., 1993; Weller and Kaas, 1985). IT neurons represent solid shape in the form of compositions of medial axial and surface components (Hung et al., 2012; Yamane et al., 2008). While the origin of these solid shape signals was heavily speculated to be V4, the issue was never tested, primarily because V4 had only been studied with planar stimuli. The results in this thesis resolve this long-standing gap by proposing that solid shape signals are incipient in V4 and, therefore, must contribute to the observed object representations in IT cortex.

4.3 Conclusion

The results presented in this thesis provide striking evidence for early visual representation of solid shape fragments in area V4 along the ventral visual pathway, an area previously thought to calculate planar image transformations. This representation was found to be explicit for fragments definable by metric shape analyses and consistent to changes in viewing conditions and different 3D cues like shading, specularity, reflectivity, refraction, and disparity. These results challenge current notions of input-output transformations in V4, raise interesting questions about the origins of these signals, and have implications for computational models of object vision, including deep network models. Two-photon imaging revealed clustering of solid and planar shape information in V4 within which neural responses were highly correlated. This clustering is the first

demonstration of neuron-resolution microarchitecture in intermediate ventral visual pathway and implies that solid shape information and planar information are perhaps processed in parallel streams in V4. The high response correlations within clusters suggest that shape tuning information may also be clustered in V4. This microarchitecture may inform hypotheses of how solid tuning emerges in V4 and the circuit mechanisms responsible for consistent, stable, explicit representation of solid object shape.

5 Appendix 1: Solid shape parameterization

The ability to describe a complex solid shape in terms of its smooth surface and its underlying structural components is an important component to the study of solid shape processing in primate vision. Just as the orientation of a curve at any point is the curve's first spatial derivative, and the curvature of the curve is its second spatial derivative, the surface orientation and curvature are first and second spatial derivatives of a surface at any point. Assuming that the surface of the object is continuous and smooth (a fair assumption given the geometry of natural object surfaces), it can be described with four parameters – principal curvatures and the orientation of the principal curvatures.

Another vital solid shape characteristic is its skeleton or medial axis – the elongated straight or curved underlying structure that typically tethers object parts to the rest of the object. The medial axis of a 2D shape is the set of points inside the shape each of which is closest to at least two points on the shape. This definition can be extended to solid shapes to describe the topological skeleton of the object. The medial axis, along with the sequence of radii along the medial axis, form the medial axis transform and completely describe a solid shape. This low-dimensional complete description is attractive for studying shape representation of natural shapes. Further, objects with discernible medial axes are abundant in natural scenes. Often these structures are a result of artifactual design (for example, tabletop and table legs) and biological growth processes (for example, mammal limbs and torso). Generation of solid shapes by growing branches from skeletal points creates potentially animate geometries used to study artificial and natural vision (Blum, 1973; Marr and Nishihara, 1978; Nevatia and Binford, 1977). Importantly, such geometries enable the study of parts-based object coding hypotheses as proposed by David Marr; psychophysical and computational experiments (Feldman and Singh, 2006; Kimia, 2003) have shown that medial axial objects have stable hierarchical representations (August et al., 1999).

Such objects are efficiently described in terms of the position, orientation, and curvature of their medial axis components. These parameters, along with knowledge of the structural connectivity of the various limbs, can produce a hugely diminished set of parameters compared to surface parameterizations.

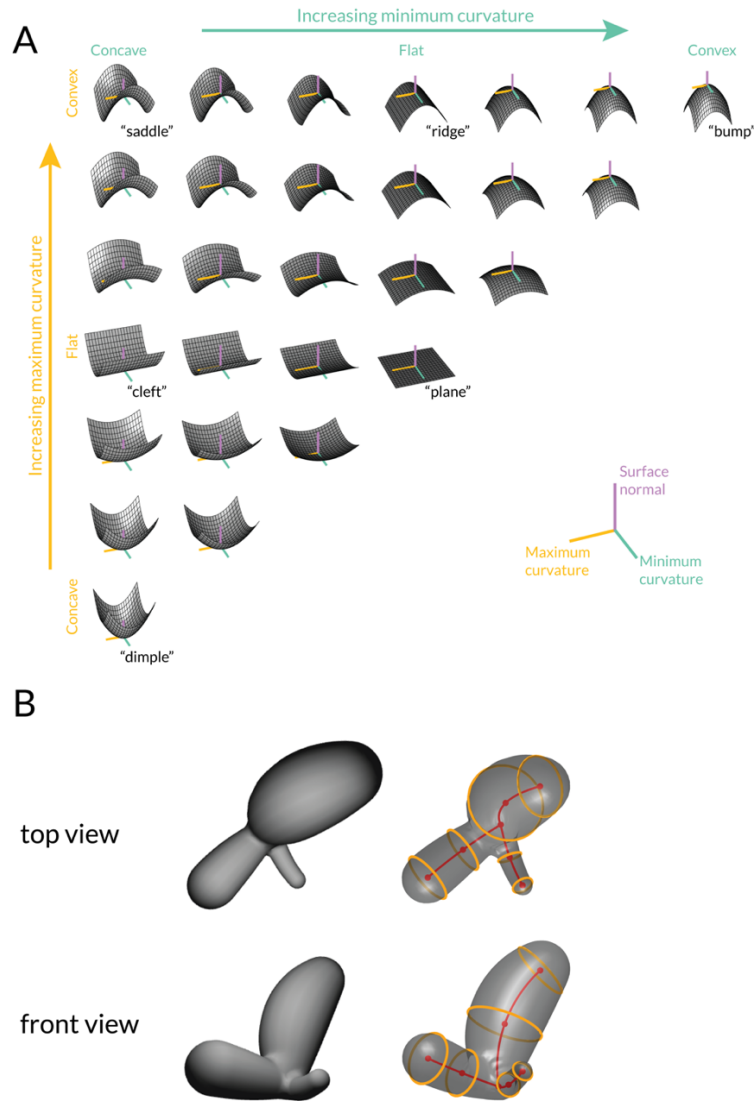


Figure 5.1: Parameterization of surface fragments and medial axial components

(A) Examples of uniformly sampled surface fragments from low to high maximum and minimum curvatures for a single surface normal direction (upwards). Purple line indicates the direction of the surface normal "out" of the surface. Cyan and yellow lines indicate the orientation of the principal curvatures. (B) Top and front views of an example stimulus used in experiments in chapter 2. The red line represents the medial axes of the shape and orange circles represent the change in radius profiles along the medial axis.

6 Appendix 2: Supplementary Figures

6.1 Supplementary figures for chapter 2

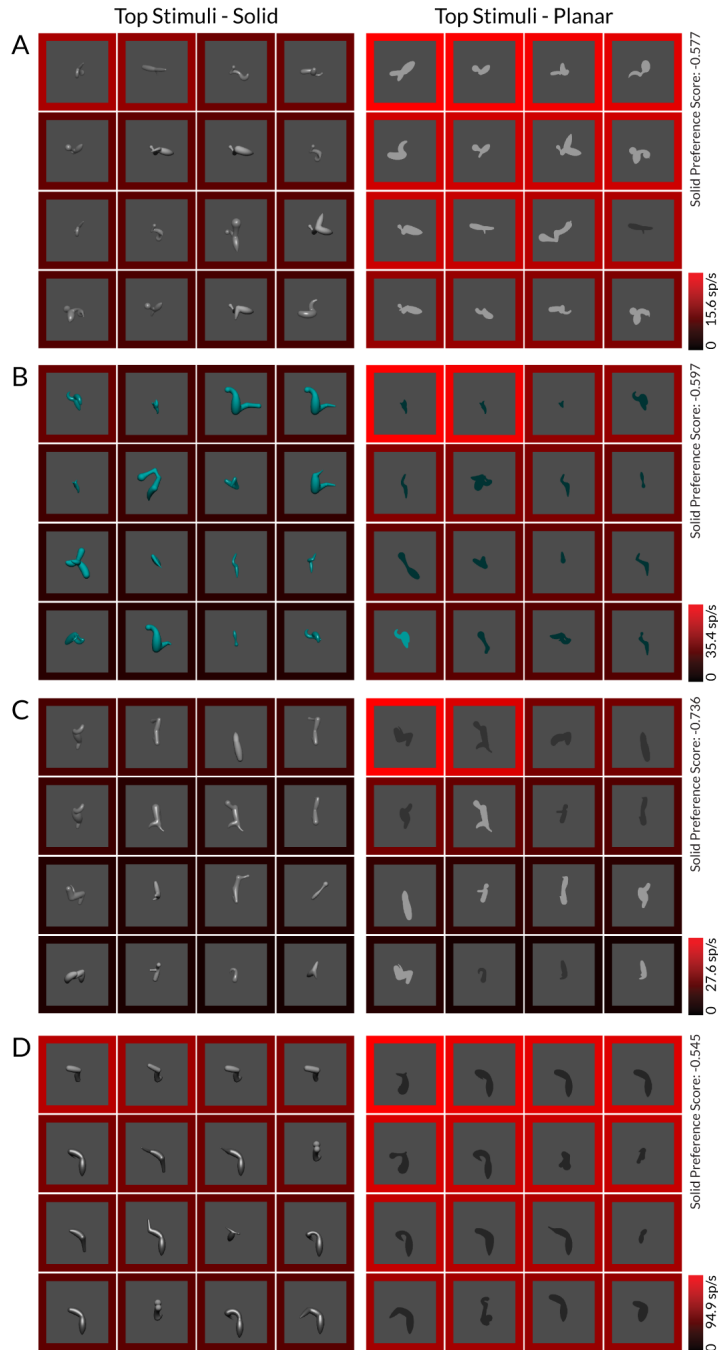


Figure 6.1: Examples of neurons highly selective for planar silhouette stimuli.

Solid shape preference score is mentioned on the right of each example.

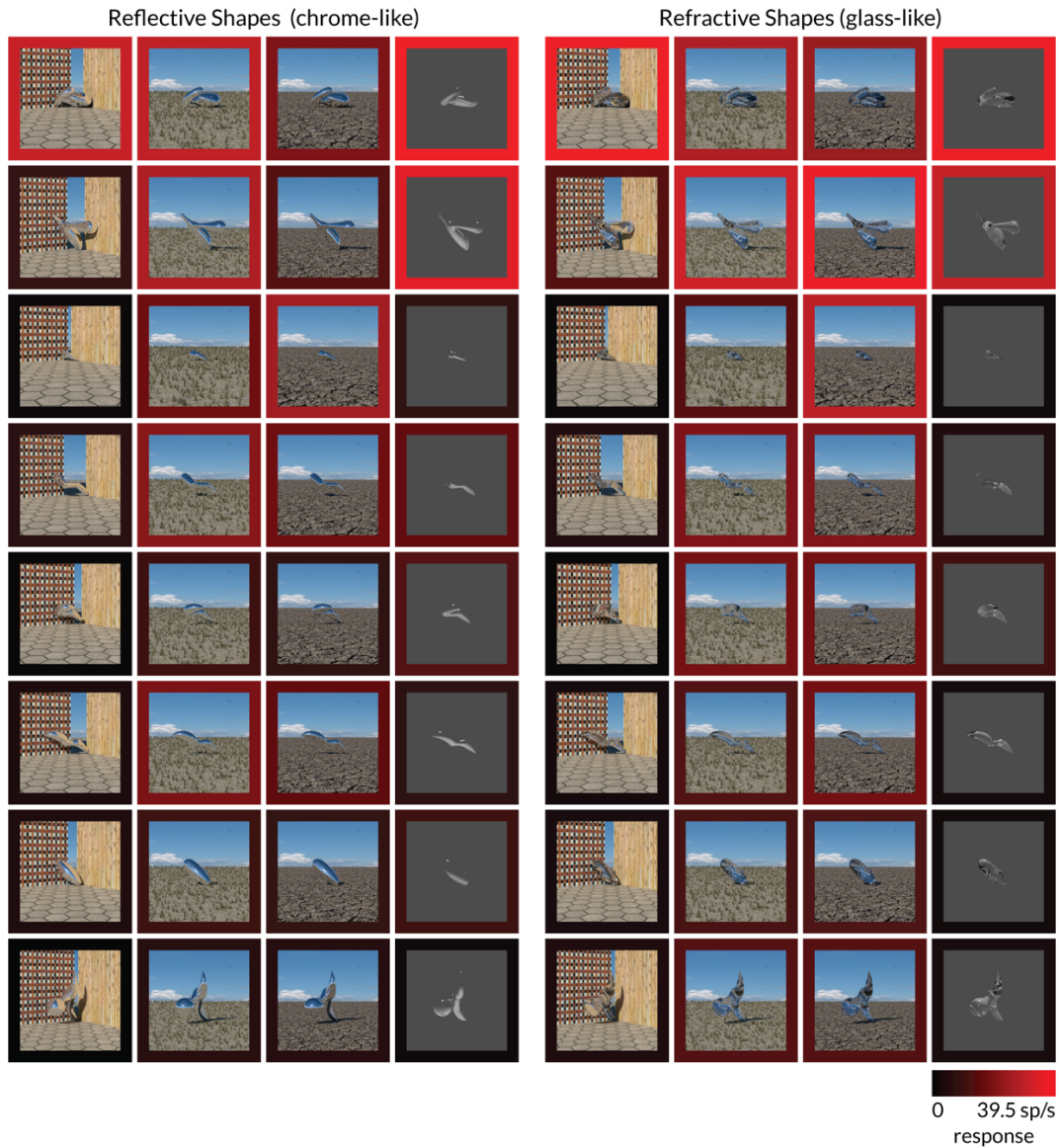


Figure 6.2: Example of a full set of stimuli in the naturalistic images post-hoc.

Full set of 40 stimuli and responses for a single V4 neuron. Responses are indicated by the image border. Eight stimuli were sampled from the high, medium, and low response ranges from the adaptive shape sampling experiment. They were rendered with two optical properties - perfect reflective (chrome-like) (left 4 columns), and perfect refractive (glass-like) (right 4 columns). They were each placed in 4 environments - closed (between 2 textured walls and a textured floor to maximize reflections/refractions), open (grassy-textured field), open (barren-textured field), and grey background. In the grey background conditions, a diffuse emissive surface was placed under the object to make the object more visible.

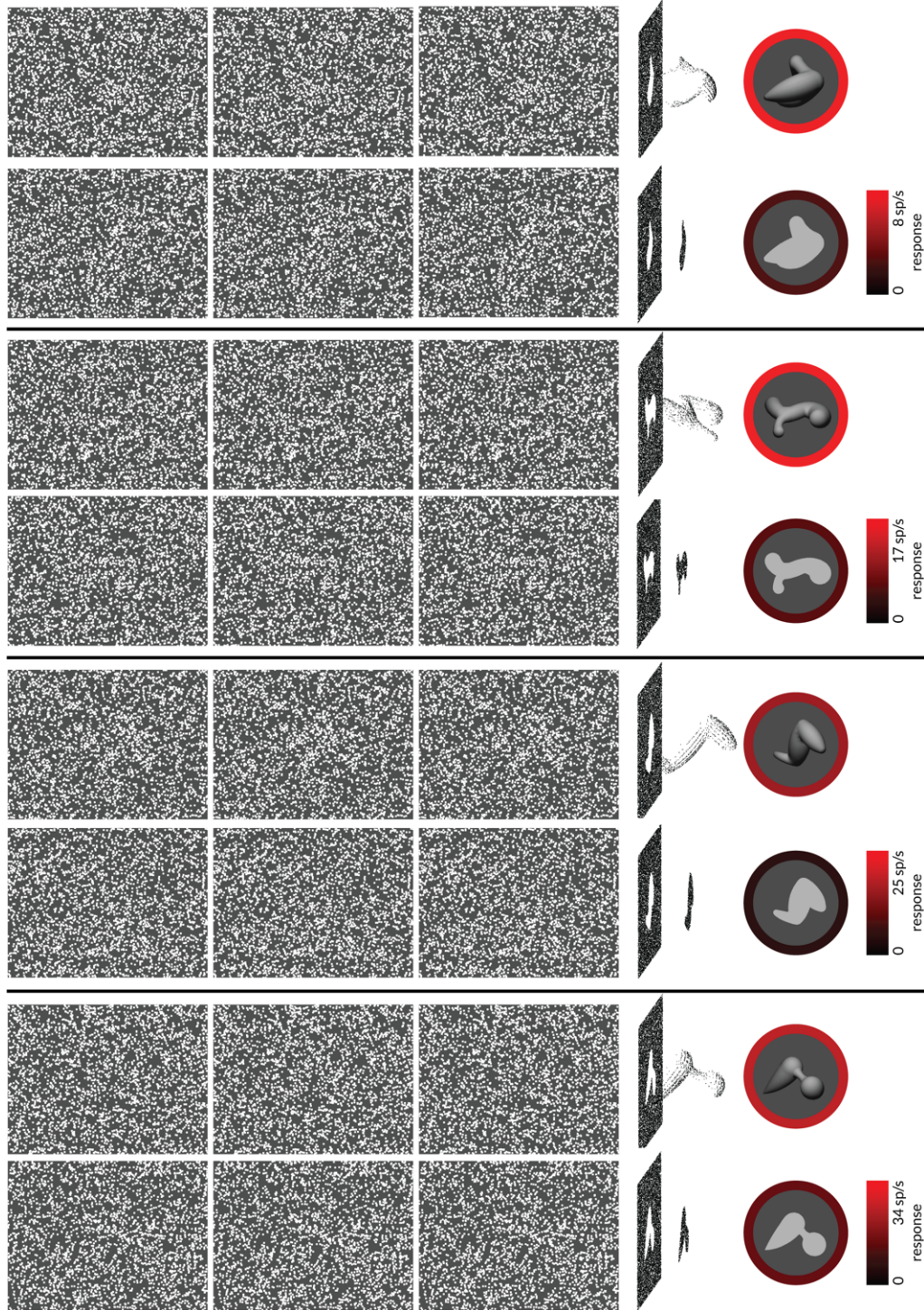


Figure 6.3: Example stereograms used in the stereogram post-hoc test.

Solid shape and flat shape random dot stereograms for the example objects in figure 2.9. For crossed fusion, use right two images. For uncrossed fusion, use left two images.

6.2 Supplementary figures for chapter 3

(These panels are for corresponding regions in panels B-F in figure 3.11)

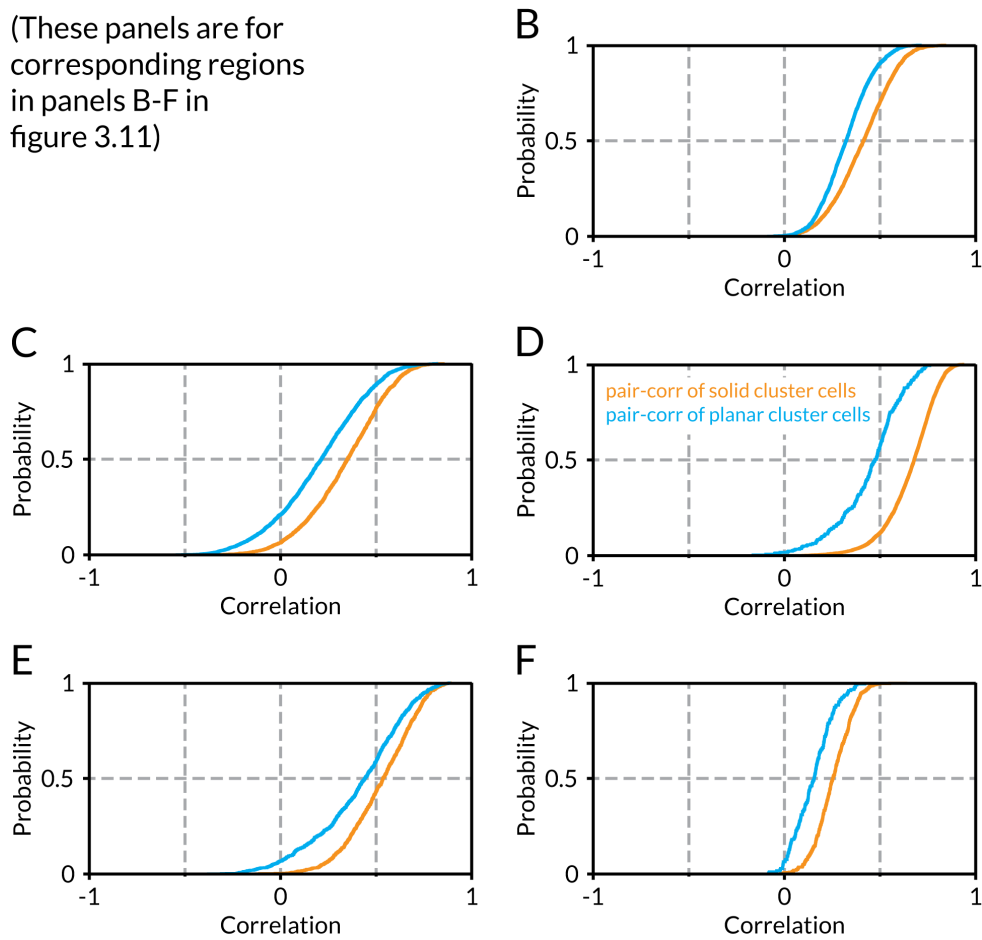


Figure 6.4: Pairwise correlation distributions for neurons within clusters.

Panels (B-F) correspond to imaging regions in panels (B-F) in figure 3.11. Each cumulative distribution of correlation values represents all pairwise correlations of neurons within a solid shape-preferring cluster (orange curves) and neurons within a planar shape-preferring cluster (cyan lines).

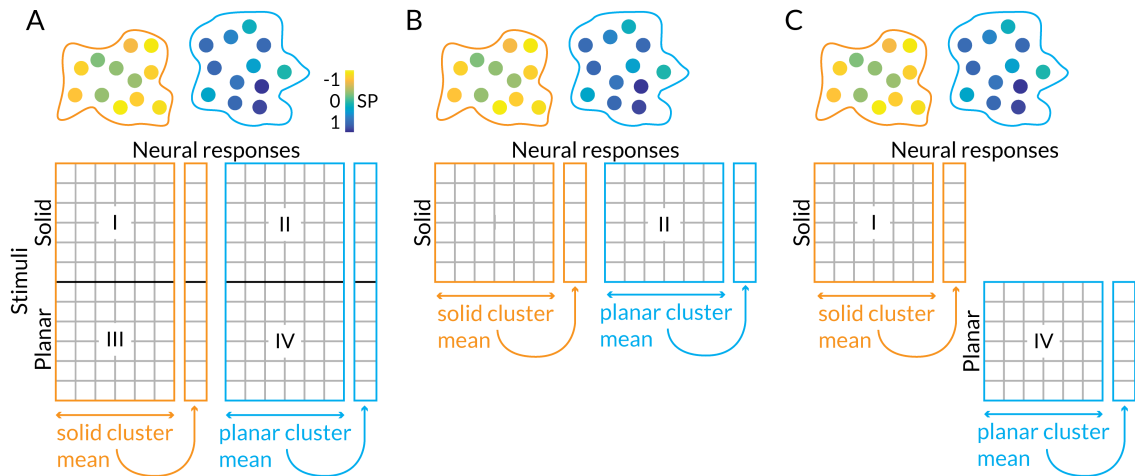


Figure 6.5: Schematics explaining the three types of correlational analysis in section 3.2.3.

Each panel depicts a matrix of stimulus responses divided into four quadrants (roman numerals) **(A)** Each neuron's response is correlated with the average response of all neurons in the solid shape-prefering cluster and the planar shape-prefering cluster across all stimuli. **(B)** Same as **(A)** except only responses from solid shape stimuli are considered. **(C)** The average response of the solid shape preferring cluster is calculated using only responses to solid shape stimuli, and of the planar shape preferring cluster using only responses from planar shape stimuli. Responses of neurons in the solid cluster only to solid shapes are correlated with the cluster averages, and responses of those in the planar cluster only to planar shapes are correlated with the cluster averages.

7 Bibliography

- Abbasi-Asl, R., Chen, Y., Bloniarz, A., Oliver, M., Willmore, B.D.B., Gallant, J.L., and Yu, B. (2018). The DeepTune framework for modeling and characterizing neurons in visual cortex area V4. *BioRxiv*.
- Allman, J.M., and Kaas, J.H. (1974). The organization of the second visual area (V II) in the owl monkey: a second order transformation of the visual hemifield. *Brain Res.* 76, 247–265.
- Arcizet, F., Jouffrais, C., and Girard, P. (2008). Natural textures classification in area V4 of the macaque monkey. *Exp. Brain Res.* 189, 109–120.
- Arcizet, F., Jouffrais, C., and Girard, P. (2009). Coding of shape from shading in area V4 of the macaque monkey. *BMC Neurosci.* 10, 140.
- Aronov, D., Reich, D.S., Mechler, F., and Victor, J.D. (2003). Neural coding of spatial phase in V1 of the macaque monkey. *J. Neurophysiol.* 89, 3304–3327.
- August, J., Siddiqi, K., and Zucker, S.W. (1999). Ligature instabilities in the perceptual organization of shape. *Computer Vision and Image Understanding* 76, 231–243.
- Baizer, J.S., Ungerleider, L.G., and Desimone, R. (1991). Organization of visual inputs to the inferior temporal and posterior parietal cortex in macaques. *J. Neurosci.* 11, 168–190.
- Barron, J.T., and Malik, J. (2012). Shape, albedo, and illumination from a single image of an unknown object. In *2012 IEEE Conference on Computer Vision and Pattern Recognition, (IEEE)*, pp. 334–341.
- Bartfeld, E., and Grinvald, A. (1992). Relationships between orientation-preference pinwheels, cytochrome oxidase blobs, and ocular-dominance columns in primate striate cortex. *Proc. Natl. Acad. Sci. USA* 89, 11905–11909.
- Bashivan, P., Kar, K., and DiCarlo, J.J. (2019). Neural population control via deep image synthesis. *Science* 364.
- Blasdel, G., and Campbell, D. (2001). Functional retinotopy of monkey visual cortex. *J. Neurosci.*

21, 8286–8301.

Blasdel, G.G., and Salama, G. (1986). Voltage-sensitive dyes reveal a modular organization in monkey striate cortex. *Nature* 321, 579–585.

Blohm, G., Keith, G.P., and Crawford, J.D. (2009). Decoding the cortical transformations for visually guided reaching in 3D space. *Cereb. Cortex* 19, 1372–1393.

Blum, H. (1973). Biological shape and visual science (part I). *J. Theor. Biol.* 38, 205–287.

Bonhoeffer, T., and Grinvald, A. (1991). Iso-orientation domains in cat visual cortex are arranged in pinwheel-like patterns. *Nature* 353, 429–431.

Born, R.T., and Tootell, R.B. (1991). Spatial frequency tuning of single units in macaque supragranular striate cortex. *Proc. Natl. Acad. Sci. USA* 88, 7066–7070.

Bradley, D.C., Chang, G.C., and Andersen, R.A. (1998). Encoding of three-dimensional structure-from-motion by primate area MT neurons. *Nature* 392, 714–717.

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

Brincat, S.L., and Connor, C.E. (2004). Underlying principles of visual shape selectivity in posterior inferotemporal cortex. *Nat. Neurosci.* 7, 880–886.

Brincat, S.L., and Connor, C.E. (2006). Dynamic shape synthesis in posterior inferotemporal cortex. *Neuron* 49, 17–24.

Burac̆as, G.T., and Albright, T.D. (1996). Contribution of area MT to perception of three-dimensional shape: a computational study. *Vision Res.* 36, 869–887.

Bushnell, B.N., Harding, P.J., Kosai, Y., Bair, W., and Pasupathy, A. (2011). Equiluminance cells in visual cortical area v4. *J. Neurosci.* 31, 12398–12412.

Cadena, S.A., Denfield, G.H., Walker, E.Y., Gatys, L.A., Tolia, A.S., Bethge, M., and Ecker, A.S. (2017). Deep convolutional models improve predictions of macaque V1 responses to natural images. *BioRxiv*.

Cadieu, C., Kouh, M., Pasupathy, A., Connor, C.E., Riesenhuber, M., and Poggio, T. (2007). A model of V4 shape selectivity and invariance. *J. Neurophysiol.* 98, 1733–1750.

- Cadieu, C.F., Hong, H., Yamins, D.L.K., Pinto, N., Ardila, D., Solomon, E.A., Majaj, N.J., and DiCarlo, J.J. (2014). Deep neural networks rival the representation of primate IT cortex for core visual object recognition. *PLoS Comput. Biol.* *10*, e1003963.
- Carlson, E.T., Rasquinha, R.J., Zhang, K., and Connor, C.E. (2011). A sparse object coding scheme in area V4. *Curr. Biol.* *21*, 288–293.
- Chance, F.S., Nelson, S.B., and Abbott, L.F. (1999). Complex cells as cortically amplified simple cells. *Nat. Neurosci.* *2*, 277–282.
- Chapman, B., Zahs, K.R., and Stryker, M.P. (1991). Relation of cortical cell orientation selectivity to alignment of receptive fields of the geniculocortical afferents that arborize within a single orientation column in ferret visual cortex. *J. Neurosci.* *11*, 1347–1358.
- Chen, G., Lu, H.D., and Roe, A.W. (2008). A map for horizontal disparity in monkey V2. *Neuron* *58*, 442–450.
- Chen, M., Li, P., Zhu, S., Han, C., Xu, H., Fang, Y., Hu, J., Roe, A.W., and Lu, H.D. (2016). An orientation map for motion boundaries in macaque V2. *Cereb. Cortex* *26*, 279–287.
- Cherniak, C. (1994). Component placement optimization in the brain. *J. Neurosci.* *14*, 2418–2427.
- Chklovskii, D.B., and Koulakov, A.A. (2000). A wire length minimization approach to ocular dominance patterns in mammalian visual cortex. *Physica A: Statistical Mechanics and Its Applications* *284*, 318–334.
- Cichy, R.M., Khosla, A., Pantazis, D., Torralba, A., and Oliva, A. (2016). Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. *Sci. Rep.* *6*, 27755.
- Connolly, M., and Van Essen, D. (1984). The representation of the visual field in parvocellular and magnocellular layers of the lateral geniculate nucleus in the macaque monkey. *J. Comp. Neurol.* *226*, 544–564.
- Connor, C.E., Brincat, S.L., and Pasupathy, A. (2007). Transformation of shape information in

the ventral pathway. *Curr. Opin. Neurobiol.* *17*, 140–147.

Conway, B.R., and Tsao, D.Y. (2006). Color architecture in alert macaque cortex revealed by fMRI. *Cereb. Cortex* *16*, 1604–1613.

Conway, B.R., Moeller, S., and Tsao, D.Y. (2007). Specialized color modules in macaque extrastriate cortex. *Neuron* *56*, 560–573.

Crair, M.C., Ruthazer, E.S., Gillespie, D.C., and Stryker, M.P. (1997). Relationship between the ocular dominance and orientation maps in visual cortex of monocularly deprived cats. *Neuron* *19*, 307–318.

Cumming, B.G., and Parker, A.J. (1999). Binocular neurons in V1 of awake monkeys are selective for absolute, not relative, disparity. *J. Neurosci.* *19*, 5602–5618.

Das, A., and Gilbert, C.D. (1997). Distortions of visuotopic map match orientation singularities in primary visual cortex. *Nature* *387*, 594–598.

David, S.V., Hayden, B.Y., and Gallant, J.L. (2006). Spectral receptive field properties explain shape selectivity in area V4. *J. Neurophysiol.* *96*, 3492–3505.

De Valois, R.L., Albrecht, D.G., and Thorell, L.G. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision Res.* *22*, 545–559.

DeAngelis, G.C., and Uka, T. (2003). Coding of horizontal disparity and velocity by MT neurons in the alert macaque. *J. Neurophysiol.* *89*, 1094–1111.

Desimone, R., and Schein, S.J. (1987). Visual properties of neurons in area V4 of the macaque: sensitivity to stimulus form. *J. Neurophysiol.* *57*, 835–868.

Desimone, R., Fleming, J., and Gross, C.G. (1980). Prestriate afferents to inferior temporal cortex: an HRP study. *Brain Res.* *184*, 41–55.

Distler, C., Boussaoud, D., Desimone, R., and Ungerleider, L.G. (1993). Cortical connections of inferior temporal area TEO in macaque monkeys. *J. Comp. Neurol.* *334*, 125–150.

Dorai, C., and Jain, A.K. (1997). Shape spectrum based view grouping and matching of 3D free-form objects. *IEEE Trans. Pattern Anal. Mach. Intell.* *19*, 1139–1145.

- Durbin, R., and Mitchison, G. (1990). A dimension reduction framework for understanding cortical maps. *Nature* 343, 644–647.
- Engel, S.A., Glover, G.H., and Wandell, B.A. (1997). Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cereb. Cortex* 7, 181–192.
- Enns, J.T., and Rensink, R.A. (1990). Sensitivity to Three-Dimensional Orientation in Visual Search. *Psychol. Sci.* 1, 323–326.
- Fang, Y., Chen, M., Xu, H., Li, P., Han, C., Hu, J., Zhu, S., Ma, H., and Lu, H.D. (2018). An Orientation Map for Disparity-Defined Edges in Area V4. *Cereb. Cortex* 29, 666–679.
- Feldman, J., and Singh, M. (2006). Bayesian estimation of the shape skeleton. *Proc. Natl. Acad. Sci. USA* 103, 18014–18019.
- Felleman, D.J., and Van Essen, D.C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* 1, 1–47.
- Ferster, D., Chung, S., and Wheat, H. (1996). Orientation selectivity of thalamic input to simple cells of cat visual cortex. *Nature* 380, 249–252.
- Fleming, R.W., Torralba, A., and Adelson, E.H. (2004). Specular reflections and the perception of shape. *J. Vis.* 4, 798–820.
- Gallant, J.L., Braun, J., and Van Essen, D.C. (1993). Selectivity for polar, hyperbolic, and Cartesian gratings in macaque visual cortex. *Science* 259, 100–103.
- Gallant, J.L., Connor, C.E., Rakshit, S., Lewis, J.W., and Van Essen, D.C. (1996). Neural responses to polar, hyperbolic, and Cartesian gratings in area V4 of the macaque monkey. *J. Neurophysiol.* 76, 2718–2739.
- Garg, A.K., Li, P., Rashid, M.S., and Callaway, E.M. (2019). Color and orientation are jointly coded and spatially organized in primate primary visual cortex. *Science* 364, 1275–1279.
- Gattass, R., Sousa, A.P., and Gross, C.G. (1988). Visuotopic organization and extent of V3 and V4 of the macaque. *J. Neurosci.* 8, 1831–1845.
- Geisler, W.S., Perry, J.S., Super, B.J., and Gallogly, D.P. (2001). Edge co-occurrence in natural

images predicts contour grouping performance. *Vision Res.* 41, 711–724.

Ghose, G.M., and Ts'o, D.Y. (1997). Form processing modules in primate area V4. *J. Neurophysiol.* 77, 2191–2196.

Glasser, M.F., Coalson, T.S., Robinson, E.C., Hacker, C.D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C.F., Jenkinson, M., et al. (2016). A multi-modal parcellation of human cerebral cortex. *Nature* 536, 171–178.

Goodale, M.A., and Milner, A.D. (1992). Separate visual pathways for perception and action. *Trends Neurosci.* 15, 20–25.

Goodhill, G.J., and Sejnowski, T.J. (1997). A unifying objective function for topographic mappings. *Neural Comput.* 9, 1291–1303.

Goris, R.L.T., Simoncelli, E.P., and Movshon, J.A. (2015). Origin and function of tuning diversity in macaque visual cortex. *Neuron* 88, 819–831.

Grinvald, A., Lieke, E., Frostig, R.D., Gilbert, C.D., and Wiesel, T.N. (1986). Functional architecture of cortex revealed by optical imaging of intrinsic signals. *Nature* 324, 361–364.

Hanazawa, A., and Komatsu, H. (2001). Influence of the direction of elemental luminance gradients on the responses of V4 cells to textured surfaces. *J. Neurosci.* 21, 4490–4497.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (IEEE), pp. 770–778.

Hegd , J., and Van Essen, D.C. (2000). Selectivity for complex shapes in primate visual area V2. *J. Neurosci.* 20, RC61.

Hegd , J., and Van Essen, D.C. (2004). Temporal dynamics of shape analysis in macaque visual area V2. *J. Neurophysiol.* 92, 3030–3042.

Hegd , J., and Van Essen, D.C. (2005). Role of primate visual area V4 in the processing of 3-D shape characteristics defined by disparity. *J. Neurophysiol.* 94, 2856–2866.

Hegd , J.A.Y., and Van Essen, D.C. (2003). Strategies of shape representation in macaque visual

area V2. *Vis Neurosci* 20, 313–328.

von der Heydt, R., Peterhans, E., and Baumgartner, G. (1984). Illusory contours and cortical neuron responses. *Science* 224, 1260–1262.

von der Heydt, R., Zhou, H., and Friedman, H.S. (2000). Representation of stereoscopic edges in monkey visual cortex. *Vision Res.* 40, 1955–1967.

Hinkle, D.A., and Connor, C.E. (2001). Disparity tuning in macaque area V4 : *NeuroReport*.

Hinkle, D.A., and Connor, C.E. (2002). Three-dimensional orientation tuning in macaque area V4. *Nat. Neurosci.* 5, 665–670.

Hinkle, D.A., and Connor, C.E. (2005). Quantitative characterization of disparity tuning in ventral pathway area V4. *J. Neurophysiol.* 94, 2726–2737.

Horton, J.C., and Adams, D.L. (2005). The cortical column: a structure without a function. *Philos. Trans. R. Soc. Lond. B, Biol. Sci.* 360, 837–862.

Huang, G., Liu, Z., Maaten, L. van der, and Weinberger, K.Q. (2017). Densely connected convolutional networks. In 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (IEEE), pp. 2261–2269.

Hubel, D.H., and Wiesel, T.N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol. (Lond.)* 160, 106–154.

Hubel, D.H., and Wiesel, T.N. (1968). Receptive fields and functional architecture of monkey striate cortex. *J. Physiol. (Lond.)* 195, 215–243.

Hung, C.-C., Carlson, E.T., and Connor, C.E. (2012). Medial axis shape coding in macaque inferotemporal cortex. *Neuron* 74, 1099–1113.

Huth, A.G., de Heer, W.A., Griffiths, T.L., Theunissen, F.E., and Gallant, J.L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature* 532, 453–458.

Issa, N.P., Trepel, C., and Stryker, M.P. (2000). Spatial frequency maps in cat visual cortex. *J. Neurosci.* 20, 8504–8514.

Janssen, P., Vogels, R., and Orban, G.A. (1999). Macaque inferior temporal neurons are selective

for disparity-defined three-dimensional shapes. *Proc. Natl. Acad. Sci. USA* *96*, 8217–8222.

Janssen, P., Vogels, R., and Orban, G.A. (2000). Three-dimensional shape coding in inferior temporal cortex. *Neuron* *27*, 385–397.

Janssen, P., Vogels, R., Liu, Y., and Orban, G.A. (2001). Macaque inferior temporal neurons are selective for three-dimensional boundaries and surfaces. *J. Neurosci.* *21*, 9419–9429.

Janssen, P., Verhoef, B.-E., and Premereur, E. (2018). Functional interactions between the macaque dorsal and ventral visual pathways during three-dimensional object vision. *Cortex* *98*, 218–227.

Ji, N., Freeman, J., and Smith, S.L. (2016). Technologies for imaging neural activity in large volumes. *Nat. Neurosci.* *19*, 1154–1164.

Jin, J., Wang, Y., Swadlow, H.A., and Alonso, J.M. (2011). Population receptive fields of ON and OFF thalamic inputs to an orientation column in visual cortex. *Nat. Neurosci.* *14*, 232–238.

Jolicoeur, P., Ullman, S., and Mackay, M. (1986). Curve tracing: a possible basic operation in the perception of spatial relations. *Mem. Cognit.* *14*, 129–140.

Justel, A., Peña, D., and Zamar, R. (1997). A multivariate Kolmogorov-Smirnov test of goodness of fit. *Stat Probab Lett* *35*, 251–259.

Kaas, J.H. (1997). Topographic maps are fundamental to sensory processing. *Brain Res. Bull.* *44*, 107–112.

Kara, P., and Boyd, J.D. (2009). A micro-architecture for binocular disparity and ocular dominance in visual cortex. *Nature* *458*, 627–631.

Khaligh-Razavi, S.-M., and Kriegeskorte, N. (2014). Deep supervised, but not unsupervised, models may explain IT cortical representation. *PLoS Comput. Biol.* *10*, e1003915.

Kim, T., Bair, W., and Pasupathy, A. (2019). Neural coding for shape and texture in macaque area V4. *J. Neurosci.* *39*, 4760–4774.

Kimia, B.B. (2003). On the role of medial geometry in human vision. *J Physiol Paris* *97*, 155–190.

- Kimmel, R. (2002). 3D Shape Reconstruction from Autostereograms and Stereo. *J. Vis. Commun. Image Represent.* *13*, 324–333.
- Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., and Broussard, C. (2007). What's new in psychtoolbox-3. *Perception*.
- Knierim, J.J., and van Essen, D.C. (1992). Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. *J. Neurophysiol.* *67*, 961–980.
- Kobatake, E., and Tanaka, K. (1994). Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *J. Neurophysiol.* *71*, 856–867.
- Koffka, K. (2013). *Principles Of Gestalt Psychology* (Routledge).
- Kosai, Y., El-Shamayleh, Y., Fyall, A.M., and Pasupathy, A. (2014). The role of visual area V4 in the discrimination of partially occluded shapes. *J. Neurosci.* *34*, 8570–8584.
- Kotake, Y., Morimoto, H., Okazaki, Y., Fujita, I., and Tamura, H. (2009). Organization of color-selective neurons in macaque visual area V4. *J. Neurophysiol.* *102*, 15–27.
- Koulakov, A.A., and Chklovskii, D.B. (2001). Orientation preference patterns in mammalian visual cortex. *Neuron* *29*, 519–527.
- Kourtzi, Z., and Kanwisher, N. (2001). Representation of perceived object shape by the human lateral occipital complex. *Science* *293*, 1506–1509.
- Kravitz, D.J., Saleem, K.S., Baker, C.I., Ungerleider, L.G., and Mishkin, M. (2013). The ventral visual pathway: an expanded neural framework for the processing of object quality. *Trends Cogn. Sci. (Regul. Ed.)* *17*, 26–49.
- Kremkow, J., Jin, J., Wang, Y., and Alonso, J.M. (2016). Principles underlying sensory map topography in primary visual cortex. *Nature* *533*, 52–57.
- Krizhevsky, A., Sutskever, I., and Hinton, G.E. (2012). ImageNet classification with deep convolutional neural networks. *Commun ACM* *60*, 84–90.
- Lafer-Sousa, R., and Conway, B.R. (2013). Parallel, multi-stage processing of colors, faces and shapes in macaque inferior temporal cortex. *Nat. Neurosci.* *16*, 1870–1878.

Landisman, C.E., and Ts'o, D.Y. (2002). Color processing in macaque striate cortex: relationships to ocular dominance, cytochrome oxidase, and orientation. *J. Neurophysiol.* 87, 3126–3137.

Lee, Y.L., and Saunders, J.A. (2011). Stereo improves 3D shape discrimination even when rich monocular shape cues are available. *J. Vis.* 11.

LeVay, S., Hubel, D.H., and Wiesel, T.N. (1975). The pattern of ocular dominance columns in macaque visual cortex revealed by a reduced silver stain. *J. Comp. Neurol.* 159, 559–576.

Levitt, J.B., and Lund, J.S. (1997). Contrast dependence of contextual effects in primate visual cortex. *Nature* 387, 73–76.

Li, M., Liu, F., Juusola, M., and Tang, S. (2014). Perceptual color map in macaque visual area V4. *J. Neurosci.* 34, 202–217.

Lim, H., Wang, Y., Xiao, Y., Hu, M., and Felleman, D.J. (2009). Organization of hue selectivity in macaque V2 thin stripes. *J. Neurophysiol.* 102, 2603–2615.

Livingstone, M., and Hubel, D. (1988). Segregation of form, color, movement, and depth: anatomy, physiology, and perception. *Science* 240, 740–749.

Livingstone, M.S., and Hubel, D.H. (1984). Anatomy and physiology of a color system in the primate visual cortex. *J. Neurosci.* 4, 309–356.

Lu, H.D., Chen, G., Tanigawa, H., and Roe, A.W. (2010). A motion direction map in macaque V2. *Neuron* 68, 1002–1013.

Malach, R., Tootell, R.B., and Malonek, D. (1994). Relationship between orientation domains, cytochrome oxidase stripes, and intrinsic horizontal connections in squirrel monkey area V2. *Cereb. Cortex* 4, 151–165.

Marr, D., and Nishihara, H.K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proc R Soc Lond, B, Biol Sci* 200, 269–294.

Merigan, W.H., and Pham, H.A. (1998). V4 lesions in macaques affect both single- and multiple-viewpoint shape discriminations. *Vis Neurosci* 15, 359–367.

- Motter, B.C. (2009). Central V4 receptive fields are scaled by the V1 cortical magnification and correspond to a constant-sized sampling of the V1 surface. *J. Neurosci.* *29*, 5749–5757.
- Mountcastle, V.B. (1957). Modality and topographic properties of single neurons of cat's somatic sensory cortex. *J. Neurophysiol.* *20*, 408–434.
- Movshon, J.A., Thompson, I.D., and Tolhurst, D.J. (1978). Spatial summation in the receptive fields of simple cells in the cat's striate cortex. *J. Physiol. (Lond.)* *283*, 53–77.
- Murray, S.O., Olshausen, B.A., and Woods, D.L. (2003). Processing shape, motion and three-dimensional shape-from-motion in the human cortex. *Cereb. Cortex* *13*, 508–516.
- Nakamura, H., Gattass, R., Desimone, R., and Ungerleider, L.G. (1993). The modular organization of projections from areas V1 and V2 to areas V4 and TEO in macaques. *J. Neurosci.* *13*, 3681–3691.
- Nauhaus, I., and Nielsen, K.J. (2014). Building maps from maps in primary visual cortex. *Curr. Opin. Neurobiol.* *24*, 1–6.
- Nauhaus, I., Nielsen, K.J., Disney, A.A., and Callaway, E.M. (2012). Orthogonal micro-organization of orientation and spatial frequency in primate primary visual cortex. *Nat. Neurosci.* *15*, 1683–1690.
- Nauhaus, I., Nielsen, K.J., and Callaway, E.M. (2016). Efficient receptive field tiling in primate V1. *Neuron* *91*, 893–904.
- Nelson, R.C., and Selinger, A. (1998). Large-scale tests of a keyed, appearance-based 3-D object recognition system. *Vision Res.* *38*, 2469–2488.
- Neri, P., Bridge, H., and Heeger, D.J. (2004). Stereoscopic processing of absolute and relative disparity in human visual cortex. *J. Neurophysiol.* *92*, 1880–1891.
- Nevatia, R., and Binford, T.O. (1977). Description and recognition of curved objects☆. *Artif Intell* *8*, 77–98.
- Norman, J.F., Todd, J.T., and Orban, G.A. (2004). Perception of three-dimensional shape from

specular highlights, deformations of shading, and other types of visual information. *Psychol. Sci.* *15*, 565–570.

Okazawa, G., Tajima, S., and Komatsu, H. (2015). Image statistics underlying natural texture selectivity of neurons in macaque V4. *Proc. Natl. Acad. Sci. USA* *112*, E351–60.

Okazawa, G., Tajima, S., and Komatsu, H. (2017). Gradual Development of Visual Texture-Selective Properties Between Macaque Areas V2 and V4. *Cereb. Cortex* *27*, 4867–4880.

Oleskiw, T.D., Pasupathy, A., and Bair, W. (2014). Spectral receptive fields do not explain tuning for boundary curvature in V4. *J. Neurophysiol.* *112*, 2114–2122.

Oleskiw, T.D., Nowack, A., and Pasupathy, A. (2018). Joint coding of shape and blur in area V4. *Nat. Commun.* *9*, 466.

Orban, G.A. (2011). The extraction of 3D shape in the visual system of human and nonhuman primates. *Annu. Rev. Neurosci.* *34*, 361–388.

Orban, G.A., Janssen, P., and Vogels, R. (2006). Extracting 3D structure from disparity. *Trends Neurosci.* *29*, 466–473.

Paik, S.-B., and Ringach, D.L. (2011). Retinal origin of orientation maps in visual cortex. *Nat. Neurosci.* *14*, 919–925.

Pasupathy, A., and Connor, C.E. (1999). Responses to contour features in macaque area V4. *J. Neurophysiol.* *82*, 2490–2502.

Pasupathy, A., and Connor, C.E. (2001). Shape representation in area V4: position-specific tuning for boundary conformation. *J. Neurophysiol.* *86*, 2505–2519.

Pasupathy, A., and Connor, C.E. (2002). Population coding of shape in area V4. *Nat. Neurosci.* *5*, 1332–1338.

Pelli, D.G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat Vis* *10*, 437–442.

Popovkina, D.V., Bair, W., and Pasupathy, A. (2019). Modelling diverse responses to filled and outline shapes in macaque V4. *J. Neurophysiol.* *121*, 1059–1077.

- Pospisil, D.A., Pasupathy, A., and Bair, W. (2018). “Artiphysiology” reveals V4-like shape tuning in a deep network trained for image classification. *Elife* 7.
- Raiguel, S., Van Hulle, M.M., Xiao, D.K., Marcar, V.L., and Orban, G.A. (1995). Shape and spatial distribution of receptive fields and antagonistic motion surrounds in the middle temporal area (V5) of the macaque. *Eur. J. Neurosci.* 7, 2064–2082.
- Ratan Murty, N.A., and Arun, S.P. (2017). Seeing a straight line on a curved surface: decoupling of patterns from surfaces by single IT neurons. *J. Neurophysiol.* 117, 104–116.
- Reid, R.C., and Alonso, J.M. (1995). Specificity of monosynaptic connections from thalamus to visual cortex. *Nature* 378, 281–284.
- Riesenhuber, M., and Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nat. Neurosci.* 2, 1019–1025.
- Roelfsema, P.R. (2006). Cortical algorithms for perceptual grouping. *Annu. Rev. Neurosci.* 29, 203–227.
- Roelfsema, P.R., and Houtkamp, R. (2011). Incremental grouping of image elements in vision. *Atten. Percept. Psychophys.* 73, 2542–2572.
- Roelfsema, P.R., and de Lange, F.P. (2016). Early visual cortex as a multiscale cognitive blackboard. *Annu. Rev. Vis. Sci.* 2, 131–151.
- Rogers, B., and Cagenello, R. (1989). Disparity curvature and the perception of three-dimensional surfaces. *Nature* 339, 135–137.
- Schiller, P.H., Finlay, B.L., and Volman, S.F. (1976). Quantitative studies of single-cell properties in monkey striate cortex. II. Orientation specificity and ocular dominance. *J. Neurophysiol.* 39, 1320–1333.
- Schrimpf, M., Kubilius, J., Hong, H., Majaj, N.J., Rajalingham, R., Issa, E.B., Kar, K., Bashivan, P., Prescott-Roy, J., Schmidt, K., et al. (2018). Brain-Score: Which Artificial Neural Network for Object Recognition is most Brain-Like? *BioRxiv*.
- Selinger, A., and Nelson, R.C. (1999). A Perceptual Grouping Hierarchy for Appearance-Based

- 3D Object Recognition. *Computer Vision and Image Understanding* 76, 83–92.
- Sereno, M.I., Dale, A.M., Reppas, J.B., Kwong, K.K., Belliveau, J.W., Brady, T.J., Rosen, B.R., and Tootell, R.B. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science* 268, 889–893.
- Shiozaki, H.M., Tanabe, S., Doi, T., and Fujita, I. (2012). Neural activity in cortical area V4 underlies fine disparity discrimination. *J. Neurosci.* 32, 3830–3841.
- Shipp, S., and Zeki, S. (1995). Segregation and convergence of specialised pathways in macaque monkey visual cortex. *J. Anat.* 187 (Pt 3), 547–562.
- Shipp, S., and Zeki, S. (2002). The functional organization of area V2, I: Specialization across stripes and layers. *Vis Neurosci* 19, 187–210.
- Shmuel, A., and Grinvald, A. (1996). Functional organization for direction of motion and its relationship to orientation maps in cat area 18. *J. Neurosci.* 16, 6945–6964.
- Siegel, R.M., and Andersen, R.A. (1988). Perception of three-dimensional structure from motion in monkey and man. *Nature* 331, 259–261.
- Silverman, M.S., Grosf, D.H., De Valois, R.L., and Elfar, S.D. (1989). Spatial-frequency organization in primate striate cortex. *Proc. Natl. Acad. Sci. USA* 86, 711–715.
- Simonyan, K., and Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. *ArXiv*.
- Sirovich, L., and Uglesich, R. (2004). The organization of orientation and spatial frequency in primary visual cortex. *Proc. Natl. Acad. Sci. USA* 101, 16941–16946.
- Skottun, B.C., De Valois, R.L., Grosf, D.H., Movshon, J.A., Albrecht, D.G., and Bonds, A.B. (1991). Classifying simple and complex cells on the basis of response modulation. *Vision Res.* 31, 1079–1086.
- Sofroniew, N.J., Flickinger, D., King, J., and Svoboda, K. (2016). A large field of view two-photon mesoscope with subcellular resolution for in vivo imaging. *Elife* 5.
- Stirman, J.N., Smith, I.T., Kudenov, M.W., and Smith, S.L. (2016). Wide field-of-view, multi-

region, two-photon imaging of neuronal activity in the mammalian brain. *Nat. Biotechnol.* *34*, 857–862.

Sun, J., and Perona, P. (1996). Early computation of shape and reflectance in the visual system. *Nature* *379*, 165–168.

Swindale, N.V., Shoham, D., Grinvald, A., Bonhoeffer, T., and Hübener, M. (2000). Visual cortex maps are optimized for uniform coverage. *Nat. Neurosci.* *3*, 822–826.

Takahata, T. (2016). What does cytochrome oxidase histochemistry represent in the visual cortex? *Front. Neuroanat.* *10*, 79.

Tanabe, S., Doi, T., Umeda, K., and Fujita, I. (2005). Disparity-tuning characteristics of neuronal responses to dynamic random-dot stereograms in macaque visual area V4. *J. Neurophysiol.* *94*, 2683–2699.

Tanigawa, H., Lu, H.D., and Roe, A.W. (2010). Functional organization for color and orientation in macaque V4. *Nat. Neurosci.* *13*, 1542–1548.

Todd, J.T. (2004). The visual perception of 3D shape. *Trends Cogn. Sci. (Regul. Ed.)* *8*, 115–121.

Todd, J.T., and Mingolla, E. (1983). Perception of surface curvature and direction of illumination from patterns of shading. *Journal of Experimental Psychology: Human Perception and Performance* *9*, 583–595.

Tootell, R.B., Silverman, M.S., Switkes, E., and De Valois, R.L. (1982). Deoxyglucose analysis of retinotopic organization in primate striate cortex. *Science* *218*, 902–904.

Tootell, R.B., Silverman, M.S., De Valois, R.L., and Jacobs, G.H. (1983). Functional organization of the second cortical visual area in primates. *Science* *220*, 737–739.

Tootell, R.B., Silverman, M.S., Hamilton, S.L., Switkes, E., and De Valois, R.L. (1988). Functional anatomy of macaque striate cortex. V. Spatial frequency. *J. Neurosci.* *8*, 1610–1624.

Touryan, J., and Mazer, J.A. (2015). Linear and non-linear properties of feature selectivity in V4 neurons. *Front. Syst. Neurosci.* *9*, 82.

Treue, S., Husain, M., and Andersen, R.A. (1991). Human perception of structure from motion.

Vision Res. *31*, 59–75.

Tsao, D.Y., Freiwald, W.A., Knutsen, T.A., Mandeville, J.B., and Tootell, R.B.H. (2003). Faces and objects in macaque cerebral cortex. *Nat. Neurosci.* *6*, 989–995.

Tsao, D.Y., Freiwald, W.A., Tootell, R.B.H., and Livingstone, M.S. (2006). A cortical region consisting entirely of face-selective cells. *Science* *311*, 670–674.

Ts'o, D.Y., and Gilbert, C.D. (1988). The organization of chromatic and spatial interactions in the primate striate cortex. *J. Neurosci.* *8*, 1712–1727.

Ts'o, D.Y., Frostig, R.D., Lieke, E.E., and Grinvald, A. (1990). Functional organization of primate visual cortex revealed by high resolution optical imaging. *Science* *249*, 417–420.

Ts'o, D.Y., Roe, A.W., and Gilbert, C.D. (2001). A hierarchy of the functional organization for color, form and disparity in primate visual area V2. *Vision Res.* *41*, 1333–1349.

Tsutsui, K.-I., Taira, M., and Sakata, H. (2005). Neural mechanisms of three-dimensional vision. *Neurosci. Res.* *51*, 221–229.

Udin, S.B., and Fawcett, J.W. (1988). Formation of topographic maps. *Annu. Rev. Neurosci.* *11*, 289–327.

Umeda, K., Tanabe, S., and Fujita, I. (2007). Representation of stereoscopic depth based on relative disparity in macaque area V4. *J. Neurophysiol.* *98*, 241–252.

Van Essen, D.C., and Maunsell, J.H. (1980). Two-dimensional maps of the cerebral cortex. *J. Comp. Neurol.* *191*, 255–281.

Van Essen, D.C., Anderson, C.H., and Felleman, D.J. (1992). Information processing in the primate visual system: an integrated systems perspective. *Science* *255*, 419–423.

Van Hooser, S.D. (2007). Similarity and diversity in visual cortex: is there a unifying theory of cortical computation? *Neuroscientist* *13*, 639–656.

Van Hooser, S.D., Heimel, J.A.F., Chung, S., Nelson, S.B., and Toth, L.J. (2005). Orientation selectivity without orientation maps in visual cortex of a highly visual mammal. *J. Neurosci.* *25*, 19–28.

- Vaziri, S., Carlson, E.T., Wang, Z., and Connor, C.E. (2014). A channel for 3D environmental shape in anterior inferotemporal cortex. *Neuron* 84, 55–62.
- Watanabe, M., Tanaka, H., Uka, T., and Fujita, I. (2002). Disparity-selective neurons in area V4 of macaque monkeys. *J. Neurophysiol.* 87, 1960–1973.
- Weliky, M., Bosking, W.H., and Fitzpatrick, D. (1996). A systematic map of direction preference in primary visual cortex. *Nature* 379, 725–728.
- Weller, R.E., and Kaas, J.H. (1985). Cortical projections of the dorsolateral visual area in owl monkeys: the prestriate relay to inferior temporal cortex. *J. Comp. Neurol.* 234, 35–59.
- Wertheimer, M. (1938). Laws of organization in perceptual forms. In *A Source Book of Gestalt Psychology*, (London, England: Kegan Paul, Trench, Trubner & Company), pp. 71–88.
- Xiao, Y., Wang, Y., and Felleman, D.J. (2003). A spatially organized representation of colour in macaque cortical area V2. *Nature* 421, 535–539.
- Yamane, Y., Carlson, E.T., Bowman, K.C., Wang, Z., and Connor, C.E. (2008). A neural code for three-dimensional object shape in macaque inferotemporal cortex. *Nat. Neurosci.* 11, 1352–1360.
- Yamins, D.L.K., Hong, H., Cadieu, C.F., Solomon, E.A., Seibert, D., and DiCarlo, J.J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proc. Natl. Acad. Sci. USA* 111, 8619–8624.
- Yau, J.M., Pasupathy, A., Brincat, S.L., and Connor, C.E. (2013). Curvature processing dynamics in macaque area V4. *Cereb. Cortex* 23, 198–209.
- Yukie, M., and Iwai, E. (1985). Laminar origin of direct projection from cortex area V1 to V4 in the rhesus monkey. *Brain Res.* 346, 383–386.
- Zeki, S. (1983). Colour coding in the cerebral cortex: the reaction of cells in monkey visual cortex to wavelengths and colours. *Neuroscience* 9, 741–765.
- Zeki, S.M. (1969). Representation of central visual fields in prestriate cortex of monkey. -
Semantic Scholar. Undefined.

Zeki, S.M. (1971). Cortical projections from two prestriate areas in the monkey. *Brain Res.* *34*, 19–35.

Zeki, S.M. (1973). Colour coding in rhesus monkey prestriate cortex. - Semantic Scholar.

Undefined.

Zhou, H., Friedman, H.S., and von der Heydt, R. (2000). Coding of border ownership in monkey visual cortex. *J. Neurosci.* *20*, 6594–6611.

Ziamba, C.M., Freeman, J., Movshon, J.A., and Simoncelli, E.P. (2016). Selectivity and tolerance for visual texture in macaque V2. *Proc. Natl. Acad. Sci. USA* *113*, E3140-9.

8 Curriculum Vitae

Ramanujan Srinath

Department of Neuroscience, Johns Hopkins University
Zanvyl Krieger Mind/Brain Institute, 3400, N Charles St, 338, Krieger Hall, Baltimore, MD-21218
ramanujan@jhmi.edu (410) 733-5258

Current position	Ph.D. candidate, Dept. of Neuroscience, Johns Hopkins University
Experience	<p>Project Assistant, Indian Institute of Science, Aug '12 – Aug '13</p> <ul style="list-style-type: none">- Lab of Dr. Supratim Ray, Center for Neuroscience, IISc- Spatial Properties of Correlations in the Amplitude of the Local Field Potential in V1- Modeling the spatial reach of ECoG electrode based on the power law distribution of LFP <p>Software Engineer, Philips Healthcare, Aug '11 – Sept '12</p> <ul style="list-style-type: none">- Log analysis algorithm design, implementation, device-side product- Lead developer: Product authorization, UI modules- Spot Awards: CAT 4.5 Delivery (Dec '11), Analysis Engine Development, UI Mockups
Internships	<p>Project Intern, Honeywell Technology Solutions Lab, Jan '11 – Jul '11</p> <ul style="list-style-type: none">- Rapid Eye DVR Adapter testing rig implementation <p>Industrial Trainee, Bharat Electronics Ltd., Jun '10 – Jul '10</p> <ul style="list-style-type: none">- Designed the test bench for MKXI SSR module for ROHINI RADAR <p>Intern, Manipal Dot Net Ltd., Sept '09 – Apr '10</p> <ul style="list-style-type: none">- Programmer for Freescale Tower with MCU CN128 <p>Industrial Trainee, Tata Communications Ltd., Jun '09 – Jul '09</p> <ul style="list-style-type: none">- Investigation of multi-service provisioning platform of metro access networks
Education	<p>PhD, Johns Hopkins University, Sept '13 – Sept '19</p> <ul style="list-style-type: none">- Department of Neuroscience, School of Medicine- Lab: Drs. Kristina J. Nielsen and Charles E. Connor- Thesis: Solid Shape Representation in Area V4 <p>Bachelor of Engineering, Manipal Institute of Technology, Aug '07 – May '11</p> <ul style="list-style-type: none">- Major: Electronics and Communication, CGPA 9.10/10 <p>12th Grade, Delhi Public School, Mathura Road, Delhi, Apr '07</p>

	<ul style="list-style-type: none"> - Majors: Physics, Chemistry, Math, English, Computer Science (C/C++), 86.2% (GPA 4.0)
	10th Grade, Cambridge School, Noida, Apr '05
	<ul style="list-style-type: none"> - Science, Math, Soc. Science, English, Sanskrit, 86.2% (GPA 4.0)
Publications	<p>Srinath, Emonds, Wang, Nielsen, Connor. (2019). Deep networks and the brain process 2D images via underlying 3D reality. (submitted)</p> <p>Srinath, Emonds, Lempel, Dunn-Weiss, Connor, Nielsen (2019). Clustering and laminar processing of 3D shape information in V4. (in preparation)</p> <p>Srinath and Ray (2014). Effect of Amplitude Correlations on Coherence in the Local Field Potential. Journal of Neurophysiology jn.00851.2013.</p>
Ongoing academic projects	<p>Deep convolution networks</p> <ul style="list-style-type: none"> - Population coding of solid shape in artificial and natural vision - Learning of solid shape discrimination from ambiguous shading cues - Coding of shape characteristics in networks with learned parts-based representations <p>Neuroscience</p> <ul style="list-style-type: none"> - Dynamics and laminar processing of solid shape representation in V4
Co-curricular projects	<p>Scholarship: Manipal University Merit Fellowship (100% tuition waiver), 2007-11</p> <p>Scholarship: Erore Educational Infotech Merit Scholarship, 2002-06</p> <p>Paper: Dupont India Challenge - 2005, Genetically Modified Foods: Hope for the Hungry or a Recipe for Disaster, Gold Certificate</p> <p>Seminar: Geodesic EEG Sensors and Brain-Computer interfacing, 2010</p> <p>Convener: OpenMic, MIT, 2010-11</p> <p>Officer: Campus Placement Committee Coordinator, 2011</p> <p>Conference: Six Model United Nation conferences across India (Deputy Secretary General and President, ECOSOC, Manipal MUN-2008)</p> <p>Award: Scholar badge for excellence in academics, 1997-2007</p> <p>Academic Consistency Award, 2007</p>
Extra-curricular projects	<p>Theatre:</p> <ul style="list-style-type: none"> - Actor, Mousetrap, JHU Barnstromers Fall Mainstage, 2015 - Lead-actor, Is he dead?, JHU Barnstromers Fall Mainstage, 2014 - Lead-actor, Noises off, JHU Barnstromers Fall Mainstage, 2013 - Lead-actor, Trouble in the Works, Comikaze, 2010 - Lead-actor, The Foreigner, 2009 - Miscellaneous acting and production assignments <p>Indian National Cadet Corp (NCC) A-certificate</p> <p>Karate: Brown belt (4th Kyu), Solo-ryu Karate Association</p>

Violin: Carnatic violin training

Football: High-school football (soccer) team

Computer skills Proficient: Java, C/C++/C# (.NET3.5 - 4), LATEX, Matlab, Python and data vis techniques

Learning: Blender, Swift (ARKit, SpriteKit, CoreML), TensorFlow

References Dept. of Neuroscience, Johns Hopkins University

- Dr. Charles. E. Connor, Professor, connor@jhu.edu

- Dr. Kristina J. Nielsen, Asst. Professor, knielse4@jhu.edu

- Dr. James Knierim, Professor, jknierim@jhu.edu

- Dr. Jay Baraban, Professor, jbaraban@jhmi.edu