Neural representation and perception of naturalistic image structure

by

Corey M. Ziemba

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Center for Neural Science

New York University

May 2016

J. Anthony Movshon

Eero P. Simoncelli

You know, the astronauts didn't read poetry. But that's changing.

Dedication

To my parents,

who taught me how to learn

Acknowledgements

Firstly, I want to thank my two advisors, Eero and Tony. The time I spent in their laboratories has been some of the happiest and most productive of my life. I have learned and continue to learn so much from both of them, and I cannot imagine having made a better decision at the outset of my graduate studies than to have chosen them as mentors.

My two closest collaborators have been Jeremy and Robbe, and the two of them have also been hugely important to me as de facto mentors. I want to thank Jeremy for inspiring me to come to NYU and showing me so much about New York, making figures, and staying up late. Robbe and I have been colleagues throughout my graduate work and he has consistently been a reliable resource for scientific support and discussion. I admire his thinking immensely, and my last few years would have been significantly diminished without our wide-ranging morning conversations. Both Robbe and Jeremy have also become two of my closest friends, and for that I consider myself incredibly lucky.

I want to thank many members of the visual neuroscience laboratory. Physiological investigation is inherently collaborative, and I've been fortunate to have benefited from the accumulated wisdom of Romesh, Luke, Najib, Brett, and Christopher. I especially want to thank Michael Gorman, who provided expert technical assistance to the collection of nearly all pieces of physiological data included in this dissertation. I want to thank Gabe for his indispensable aid in freeing me from some of my data collection duties, without which I would likely not have been able to complete this dissertation.

I want to thank many members of the laboratory for computational vision

including Deep, Umesh, Neil, Yan, James, Rich, Josh, Alex, Olivier, and Rob. Together they, along with many others, created a fantastic environment that I have always felt very fortunate to be a part of. I want especially to acknowledge Andrew for his support, insight, and his presence with me during every step of our graduate school careers.

Special thanks to the members of my committee: Mike Hawken, Roozbeh Kiani, and Bruce Cumming. I also want to thank Lynne Kiorpes, and Mike Landy, who were important collaborators on projects that I could not include in this dissertation.

Finally, I want to thank my friends and family. Thanks to my parents and my sister for their support, and for helping me escape my studies and the city a couple times a year. Max, Ben, and Carlos are old friends, but they have stayed my closest in New York. Thanks to Will, Conor, Joe, and Ari for coming to visit me, and for their hospitality when I visited them. And thanks to my fellow graduate student Silvia, for being the best thing to come out of getting a PhD.

Preface

Chapters 2 and 3 were the result of a close collaboration with fellow graduate student Jeremy Freeman in the labs of Eero Simoncelli and Tony Movshon. David Heeger was also involved in developing parts of chapter 2. Several portions of both chapters were also included in Jeremy's dissertation [58], and most of this work has been published [61, 260]. Chapter 4 represents an ongoing follow-up to the first two chapters, also in collaboration with Jeremy. Chapter 5 presents developing work from a project in collaboration with Robbe Goris.

Abstract

The perception of complex visual patterns emerges from neuronal activity in a cascade of areas in the cerebral cortex. As information propagates along this hierarchy, neuronal responses become more selective for particular features of natural images and more tolerant to image transformations that preserve those features. Despite this generally accepted framework for visual processing, we lack satisfying descriptions of the neural computations supporting visual representations at intermediate stages of this pathway. The role of the second visual area (V2) in pattern vision has been particularly enigmatic, partly because no simple response properties robustly distinguish V2 neurons from their inputs in primary visual cortex (V1). Previous approaches to midlevel vision and V2 have employed intuitions about features of intermediate complexity in natural images, or have attempted to build models to predict neuronal responses to arbitrary natural images. Here, we employ a new approach, constructing targeted, naturalistic stimuli by building on insights from models of the V1 inputs into V2, the statistics of natural images, and perception. We found that V2 neurons, but not V1 neurons, responded more vigorously to naturalistic texture stimuli than to control stimuli that lacked the statistical dependencies found in natural images. V2 responses were also linked to both the detection and discrimination of naturalistic structure in perceptual experiments in humans and nonhuman primates. Together, our results suggest a specific role for V2 in visual perception, and a framework for approaching midlevel computations in hierarchical sensory transformations.

Contents

List of Figures

Chapter 1

Introduction: An approach to mid-level vision

Vision is supported by a cascade of dozens of different brain areas working in concert to produce our experience of the visual world [121]. The study of vision encompasses the investigation of our visual behaviors as well as their neural basis. As primarily visual animals, humans have a vast repertoire of highly developed and complex visual capabilities, and devote an exceedingly large proportion of their brains to processing visual information. This neural specialization is shared with our primate ancestors as well as many other species who also rely on vision as a primary and essential sense [246, 52, 181]. No visual capability is as striking as our talent for quick object recognition and scene understanding [46]. The remarkable nature of this capacity can be obscured by its effortlessness. But the complexity of this fundamental component of human vision is underscored by the fact that until only the last handful of years, no artificial systems could even approach the performance of human observers at recognizing objects, despite attempts that stretch back for over a half century.

One goal of visual neuroscience is to understand the contribution of each differ-

ent area of the brain in supporting our visual capabilities [18]. We often conceive of the visual system as a branching hierarchy, with each area of the brain processing particular aspects of visual information (Fig. 1.1, [52]). Representations at different processing stages become increasingly complex as information flows from the retina to deeper structures in the brain [179]. This hierarchy has also been grossly divided into two different visual streams, each comprising multiple visual cortical areas [231, 230]. The dorsal visual stream is often described as processing vision for action and the ventral visual stream as processing vision for recognition (Fig. 1.1).

A major approach over the last many decades has been to investigate the properties of single and populations of neurons at each stage of this visual hierarchy. The goal is to elucidate what different neurons represent about the visual world, and how they can accomplish this through computations on their inputs [87, 32]. This has proved a difficult endeavor beyond the earliest stages because our visual behaviors are complex, and the dimensionality of the visual input is enormous. With current technology, we can record from one to hundreds of neurons simultaneously and expose them to tens of thousands of stimuli at the very most. The space of possible visual images is effectively infinite, however, and this curse of dimensionality means that choices must be made about the subset of stimuli to use in characterizing the visual selectivity of neurons. Those choices can have a profound impact on the conclusions that we draw from physiological experiments.

What principles can guide the choice of stimuli? Often the choice has come down to intuitions about what is important in the visual world. As will be discussed further, this has worked well at both the top and bottom of the visual hierarchy. This intuitive approach has given us a broad understanding of what

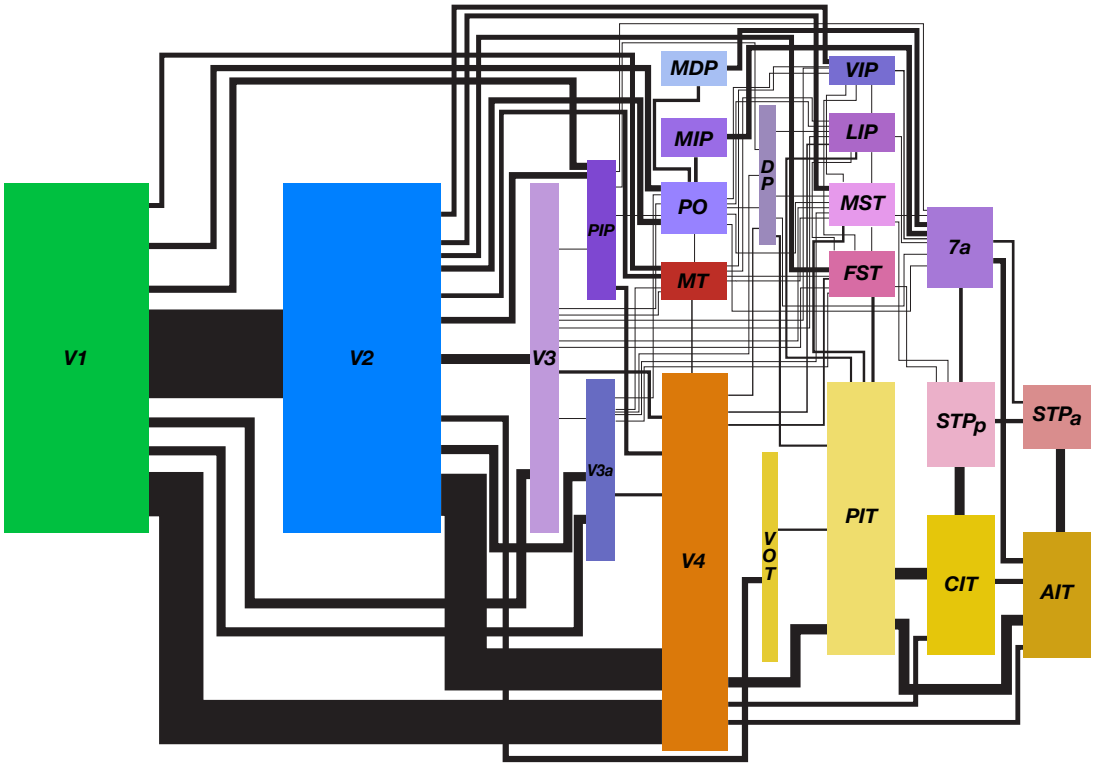

Figure 1.1: The primate visual system. Each visual cortical area in the macaque monkey is represented as a colored box proportional in size to the cortical surface area. Connections are represented as lines whose thickness represents the number of fibers in the connection. Dorsal stream areas are represented on the top right, and ventral stream areas on the bottom right. From [245], adapted from [52]

kinds of visual information are represented at many stages of the visual system. However, in order to claim a true understanding of the visual system, we need models that can account for the responses of single neurons to arbitrary stimuli. Additionally, we need those models to be interpretable to gain insight into the perceptual consequences of particular neural representations. We have attained something approaching this success in only a handful of areas of the visual brain [32], and have yet to achieve it fully even in the retina [56, 178].

Understanding visual neurons has been particularly difficult at the middle levels of the ventral stream of the visual pathway, thought to underlie vision for recognition. Here, our intuitions about the building blocks of the visual world have largely failed us. In this thesis, my collaborators and I have attempted to outline a new approach, guided by perceptual observations, the statistics of natural images, and a model-based understanding of visual representations at lower levels of the visual system.

1.1 Transformations in early vision

Vision begins in the retina, and the visual selectivity of single neurons was first investigated there. The responses of intrinsically light sensitive photoreceptors in the retina are transformed into the receptive fields of retinal ganglion cells (RGCs), the output cells of the retina. Single RGCs respond to increments (ON-centered) or decrements (OFF-centered) of light at a particular location in the visual field. In the primate, most receptive fields in the retina and its major target, the lateral geniculate nuclues (LGN), are circularly symmetric and exhibit antagonistic surround regions driven by the opposite polarity of light [249]. This generic subcortical receptive field is radically altered once visual signals reach the

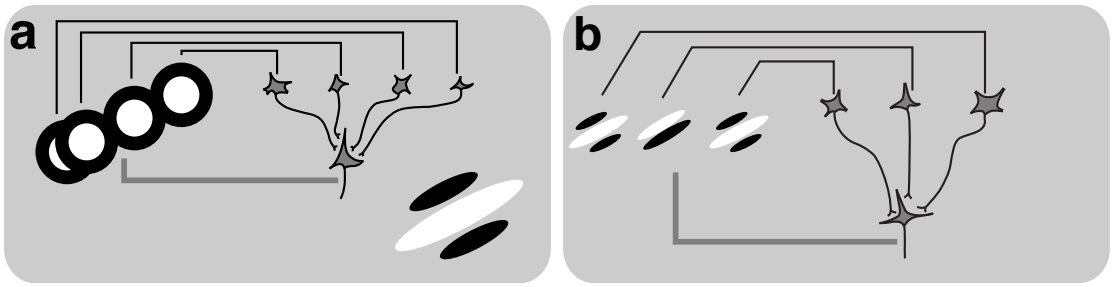

Figure 1.2: Construction of simple and complex receptive fields in V1. (a) The hypothesized scheme for constructing oriented, V1 simple cell receptive fields by combining unoriented LGN receptive fields. (b) The hypothesized scheme for constructing complex cells. Simple cell receptive fields of differing position but the same orientation are combined to form an oriented, but spatially invariant complex cell receptive field. Adapted from [101]

cerebral cortex.

Hubel and Wiesel discovered a major reorganization in the cortical sensitivity to patterns of light. They found that single neurons in the primary visual cortex (V1) are selective for the orientation of luminance edges ([101, 102], Fig. 1.2a). Mapping receptive fields using simple spots, bars, and edges of light, they discovered that some V1 cells (referred to as a "simple" cells) were characterized by nonoverlapping, elongated subregions selective for increments and decrements of light. The organization of these elongated subregions endowed the overall receptive field with selectivity for the orientation of edges. This finding uncovered a wholly new form of visual selectivity than is present in the LGN inputs into V1. Hubel and Wiesel suggested this transformation could be accomplished by summing the responses of LGN neurons whose receptive fields were aligned along the axis of preferred orientation ([101], Fig. 1.2a).

Hubel and Wiesel also discovered "complex" cells that contained no subregions with specific ON or OFF selectivity. These neurons were still selective for orientation, but invariant to the polarity, and the exact position of the oriented edge

([101], Fig. 1.2b). By analogy to the construction of simple cells from combinations of LGN receptive fields, Hubel and Wiesel proposed that complex cells were constructed by combining multiple simple cells with the same orientation preference but different positions. The multiple overlapping ON and OFF subregions of the simple cells generate invariance to position and polarity but retain selectivity to orientation.

The discovery of the emergence of orientation selectivity in cortex, and the proposed hierarchical transformation of simple cell receptive fields into complex cell receptive fields has been hugely influential. Although the proposed mechanisms for receptive field construction were not necessarily correct at a quantitative level, The basic intuition was sound. And in the years since these discoveries, this conceptual model of sensory transformation has been made more concrete by formalizing it within an explicitly computational framework. Linear systems analysis allowed researchers to test the validity of intuitions about how the selectivity of these early visual neurons arose.

The first attempts at applying a more rigorous, model-based understanding of receptive fields were made in the retina. A simple linear receptive field mechanism followed by a static nonlinearity could account very well for the receptive fields of many RGCs [47]. However, some RGCs exhibit profound nonlinearities in their summation properties over space [47]. Still, the nonlinearity of these neurons could be explained by the action of many smaller, linear subunits whose outputs are combined following a rectification [97, 234]. This subunit idea was a powerful one that faciliated the description of highly complex selectivity within a computational framework that was eminently workable [59]. Along with the concept of linear filtering, static nonlinearities, and normalization, subunit architecture has become

part of a toolkit of canonical neural computations that appear to recur throughout the visual system as well as other sensory and nonsensory systems in the brain [87, 33, 139].

Over many years, our picture of the retina has grown increasingly complex [56, 135, 14]. Highly specialized, even cortex-like, receptive fields have been discovered in the retinas of many animals, including the rabbit and the mouse [19, 25, 215, 192]. However, these complicated retinal receptive fields have yet to be discovered in many higher mammals like the primate, and even if they are present must be hugely outnumbered by simpler receptive field types. Although there is likely much more retinal complexity yet to be discovered, models of V1 built upon a linear, pixel-like input have been very successful.

Just as we have a standard model for retinal receptive fields, a standard model of V1 has also developed. The responses of simple cells can be captured with a linear filter and nonlinearity (Fig. 1.3, [142]). The structure of the linear filter distinguishes subcortical from V1 responses here. To model the nonlinear behavior of V1 complex cells, the subunit architecture from the retina can be modified such that simple cells constitute the subunits. In this framework, multiple simple cell-like linear filters are applied to an image, rectified and then summed [141]. For most purposes, the responses of V1 complex cells can be summarized using just two subunits in quadrature phase ([4, 227], Fig. 1.3). Capturing the full complexity of V1 responses requires additional subunits, and can be done effectively with an efficient convolutional subunit architecture [187, 130, 240]. While such models can account for a great deal of the variance in V1 responses, there are many additional known (and also likely unknown) nonlinear properties of V1 neurons, such as adaptation and modulatory influences from beyond the classical receptive field.

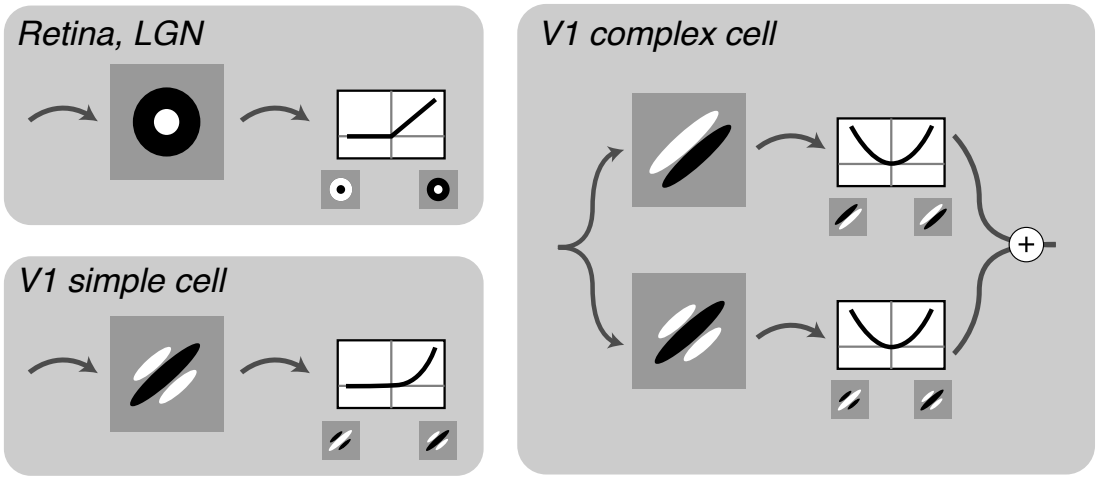

Figure 1.3: Commonly used models to summarize the response selectivity of neurons in the retina, LGN, and V1. Each model consists of one or more linear receptive fields, followed by a pointwise nonlinearity such as rectification or squaring.

Several of these can be reproduced through modifications to the basic modeling framework; for example, by adding normalization, adaptation, or other forms of gain control [85, 31, 32, 186, 134, 38].

It is useful to step back and ask what success at characterizing the basic selectivity of the early visual system tells us about the prospects for understanding neurons in later areas. Although receptive fields in the retina, LGN, and V1 were first characterized using simple spots and edges of light, that does not mean the purpose of these neurons is to detect the presence of such stimuli. Indeed, it is precisely moving away from thinking of V1 neurons as edge or line detectors that has allowed their thorough quantitative characterization. Adelson and Bergen referred to this as thinking of the early visual system as measuring "stuff" rather than detecting "things" [6]. Many aspects of receptive fields make intuitive sense within this framework, as they appear to be measuring structured changes in visual substances along dimensions such as space, time, wavelength, or depth. A V1 sim-

ple cell can be thought of as an oriented spatial derivative, a directionally selective V1 cell as a derivative oriented in space-time. This powerful conceptualization of early vision has been difficult to carry downstream, because we have no concept for what the "stuff" of midlevel vision might be, at least in the domain of form vision.

1.2 Approaching midlevel vision

There has been a great deal of success in characterizing midlevel vision, it is only that it has mostly been accomplished in visual areas that are thought to be part of the dorsal stream [231]. How should one approach the study of vision outside of V1? A good first step is to identify how visual signals might be elaborated in the extrastriate area under study. One area where this first step was immediately successful is area MT (V5), where nearly all neurons are selective for the direction of visual motion.

Selectivity for direction of motion first arises in V1. Hubel and Wiesel, along with their discovery of the emergence of orientation selectivity in V1, also discovered that a subset of V1 neurons are additionally sensitive to the direction of motion of an oriented edge [101, 102]. However, the pervasiveness of direction selectivity in MT distinguishes its neurons from those in V1 quite well (Fig. 1.4a, [140, 22]). If one were to play a hypothetical physiologists game to identify the cortical origin of a single neuron on the basis of one a single response property, one would be correct over 80% of the time by using direction selectivity to distinguish V1 from MT (Fig. 1.4a). In addition to the increased selectivity for motion direction in MT, it was also discovered that several MT cells exhibited further complexity that was completely absent in V1. Some MT neurons could signal the

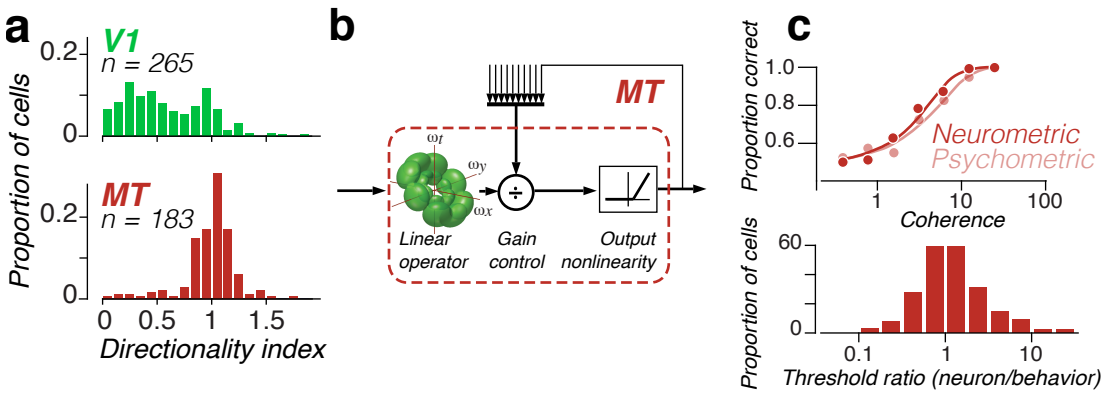

Figure 1.4: Characterization of area MT. (a) Distribution of direction selectivity for single neurons in V1 (top) and MT (bottom). The directional index was computed as $1 - n/p$, where p is the response to the preferred direction and n is the response to the direction 180° away. Adapted from [140]. (b) Computational framework for MT. The input to a single MT neuron is a simulated V1 population response. Adapted from [87]. (c) Comparing neuronal sensitivity with behavioral sensitivity. Comparing sensitivity to coherent motion of a single MT neuron with the simultaneously measured behavior of a nonhuman primate observer. Distribution of the ratio of neuronal threshold to psychophysical threshold across 216 sessions. Adapted from [27].

direction of motion of a moving visual pattern invariant to the spatial structure of the pattern [144].

Once the selectivity of the inputs has been differentiated from that of neurons in the target area, one must understand how this elaboration can arise through computation performed on the inputs. How could V1 inputs be combined to form pattern cells in MT tuned to the velocity of a visual pattern and relatively invariant to the structure of that visual pattern [144]? The ultimate goal is to characterize the computation well enough to predict the spikes of individual neurons to arbitrary stimuli. However, this can be very difficult for a variety of reasons, and deriving a qualitative match to neural data can be challenging enough. In the case of MT, there was a theoretical understanding of how MT neurons might acquire their selectivity long before that model was instantiated in a way that could predict

the firing rate of single neurons (Fig. 1.4b, [87, 204]). Subsequently, hierarchical models for MT have been constructed to operate on a simulated V1 output and fit to individual MT neurons [185, 156, 257].

Finally, one can ask how an area supports perception of the visual features represented by its neurons. There are many indirect ways to compare the similarity of a perceptual representation to a given level of neuronal representation [117, 165]. When possible, the ideal experiment is to record from or perturb a population of neurons while a subject is engaged in perceptual behavior thought to depend on those neurons. In these types of experiments, first carried out in area MT [149, 27], investigators have consistently found a close relationship between the sensitivity of single sensory neurons and the perceptual sensitivity of the entire organism (Fig. 1.4c, [27, 36, 152, 174, 229, 155]. The tightness of this relationship can be affected by several factors, including timing [39] and optimization of the stimuli for the neuron under study [80, 79, 175, 198]. In most cases, at the very least, the most sensitive neurons in a sample approach the sensitivity of the animal [165, 151]. MT has been tied to perceptual behavior in several other ways as well (but see [84]): Lesioning MT causes a specific and profound impairment of motion perception [148], microstimulation of MT biases motion judgments [188, 189], and the responses of single MT neurons to the same stimulus predict behavior on a trial-to-trial basis [26].

Thus, to understand a midlevel visual area we must differentiate the selectivity of its neurons from that of its inputs, develop a functional understanding of the computations its neurons perform to achieve this selectivity, and link this selectivity to perceptual behaviors. Arguably, MT is the only area where all three have been adequately explored. There are many areas of the brain beyond V1 where

one or two of these tests have been passed for a specific stimulus domain. There are many more areas for which we really have not succeeded fully in any of these endeavors, the most surprising and enigmatic of which is the second visual area, or V2.

1.3 The second visual area

The role of V2 has been particularly inscrutable, despite its apparently substantial role in many aspects of vision. V2 is the largest extrastriate visual cortical area in primates [226], and the most prominent recipient of projections from V1 (Fig. 1.1, [52, 121, 190]). It makes strong projections to areas in both the dorsal and ventral stream [199, 44, 66], and its neurons exhibit selectivity to a wide range of elemental visual properties [123, 68, 210].

Anatomical staining of V2 in some species of primate reveals a tripartite, stripe-like compartmental structure [226, 225, 210]. Tracers indicate that different compartments receive distinct inputs from different anatomically defined subpopulations of V1 [209, 49, 50]. Such studies suggest that some V2 neurons, those in the "thick" stripes, may be more a part of the dorsal visual pathway and project to MT, whereas other neurons, those in the "thin" and "pale" stripes, may be more involved in ventral stream processing [44, 199]. While some studies have found functional differences between neurons in different stripe types [100, 200, 131, 129, 68], others have found little to differentiate them [123, 201]. The functional significance of these compartments, like much else in V2, remains to be fully elucidated [98].

Mirroring the greater success in understanding the dorsal stream, the role of V2 has been clearer within stimulus domains associated with this visual pathway. In

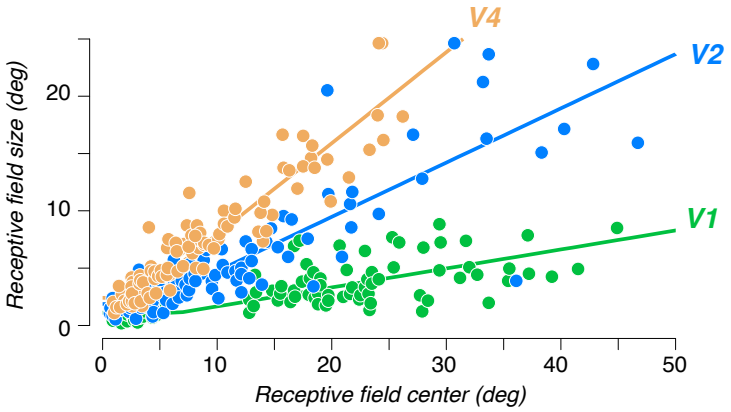

Figure 1.5: Receptive field size scales with eccentricity at different rates in V1, V2, and V4. V2 receptive fields are roughly twice the diameter of V1 at all eccentricities. Adapted from [60], data from [64, 65]

particular, the representation of disparity has been well characterized. V2 neurons can be strongly differentiated from their V1 inputs on the basis of selectivity for absolute or relative disparity cues [220, 24]. The transformation from the representation of absolute disparity by V1 neurons to that of relative disparity by V2 neurons can be explained by using computational models that utilize a subunit architecture [220, 23, 217]. Finally, V2 neurons, and not V1 neurons, have been shown to predict behavioral judgments of disparity [152, 153]. These representational properties in V2 are vital for further processing in the dorsal stream [171, 213], and may also provide important input to ventral stream structures. However, establishing a role for V2 in ventral stream processing of pattern and form information has proven much more difficult.

Anatomically, V2 appears to be central to processing in the ventral stream pathway. V2 makes a large projection to area V4 [66, 232], the largest midlevel ventral stream area (Fig. 1.1, [180]), as well as areas further up the ventral hierarchy [145]. Lesioning V2 in macaque monkeys leaves many aspects of vision intact, but impairs performance on certain tasks such as visual search within textured displays which suggest a role in midlevel form vision [138].

The responses of V2 neurons depend on feedforward input from V1 [190, 210],

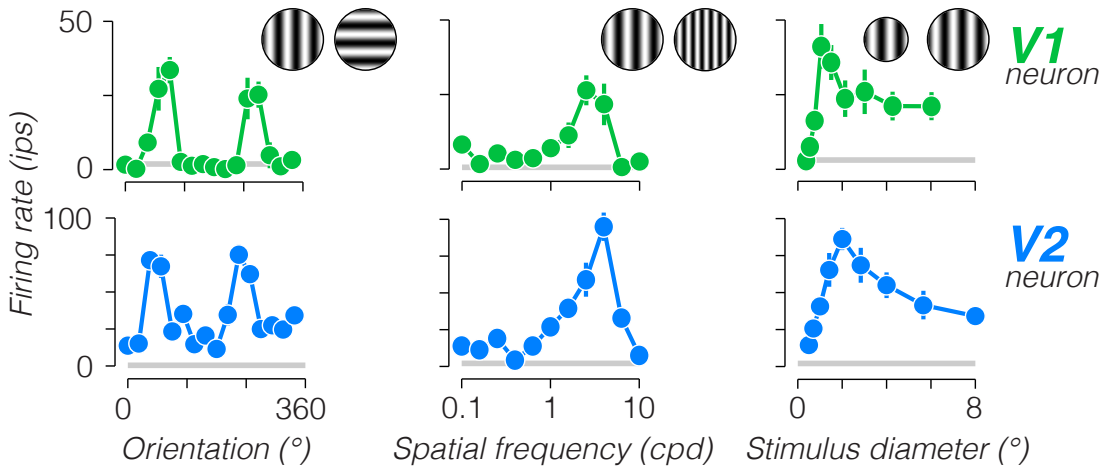

Figure 1.6: Similarity of tuning in V1 and V2. Three tuning curves gathered for drifting grating stimuli that varied in orientation (left), spatial frequency (middle) and size (right) for single V1 (top) and V2 (bottom) neurons recorded from an anesthetized macaque monkey. Error bars represent s.e.m. across repetitions.

and their receptive fields are roughly twice the diameter of those in V1 (Fig. 1.5, [64, 201]). Just as receptive fields in V2 are larger than those in V1, receptive fields in V4 are larger than those in V2 (Fig. 1.5). This progression of receptive field sizes in the ventral stream suggests that each stage combines multiple receptive fields from previous stages, and is thought to be mirrored by a progression in receptive field complexity. Unlike MT, which receives V1 input from a functionally limited subset of directionally selective complex cells [140], V2 receives input from the full diversity of receptive fields that exist in V1 [195]. V2 is most likely combining and elaborating signals from V1 to encode image features that V1 does not, but simply examining the properties of its inputs provides little insight.

Examining the selectivity of V2 neurons using the kinds of stimuli often used to characterize V1 reveal few noteworthy differences (Fig. 1.6). V2 neurons are selective for orientation and spatial frequency content, similar to V1 neurons [57, 123, 75, 51, 121]. V2 neurons also have similar proportions of selectivity for di-

rection of motion and chromatic properties as those in V1 [51, 121, 114, 163]. Many studies have subsequently resorted to more elaborate stimuli in attempting to reveal differences between V1 and V2 neurons.

Hubel and Wiesel made some of the first attempts to try to understand how extrastriate areas might elaborate on the receptive fields they encountered in V1. Similar to their formulation of simple and complex cells in V1, they posited that increased receptive field complexity in V2 neurons could come from simple combinations of their inputs. They proposed that V2 neurons might combine V1 neurons of differing preferred orientation to create selectivity for corners of a specific angle [103]. This idea reveals an inclination toward thinking of visual neurons functioning as detectors of specific "things" in the environment that seem intuitively important for building complex visual representations. Their intuition that, following the emergence of orientation selectivity in V1, the next logical step in vision is curvature or angle selectivity has been reflected in many studies over subsequent decades [89, 91, 132, 106, 11, 105]. Nonetheless, such properties have largely failed to reliably differentiate V2 neurons from their V1 inputs (Fig. 1.7).

Another related approach suggests that V2 neurons are tuned for orientation like those in V1, but can signal the orientation of features more complex than luminance. Oriented edges can be signaled by luminance, but also with so-called anomolous contours or sharp transitions in texture. Using such stimuli has revealed some interesting aspects of visual representation in V1 and V2, but only modestly differentiated their response properties (Fig. 1.7, [169, 120, 219, 196, 83, 125]).

Von der Heydt and colleagues have extensively investigated the selectivity of V2 neurons for the figure-ground structure of an oriented edge [259]. Compared with V1, a larger subset of V2 neurons seem to be sensitive to the "border ownership"

of an oriented edge (Fig. 1.7). However, these signals rely on global contextual information from far outside the receptive field of V2 neurons, may depend on top-down visual attention [176, 48], can persist when contextual cues are removed [159], and even remap across the visual field [160]. Thus, although apparent in the responses of V2 neurons, it remains to be seen whether this signal represents computations being carried out by V2 neurons or rather feedback from higher areas in the ventral stream [94].

A problematic feature of such approaches to understanding midlevel vision is the inability to explain or even engage with other potential forms of selectivity. Characterizing V2 responses within a handpicked, parametric feature space such as those of curvature, texture-defined edges, or figure-ground organization does not allow us to predict how V2 neurons would respond to other classes of stimuli, including natural images. This presents a limitation, as validating models using natural images represents an important benchmark for gauging our understanding of neural representations [186]. The visual system evolved to represent natural images, and using highly complex or natural stimuli has facilitated our understanding of areas deep in the ventral stream such as inferotemporal cortex (IT). Here, our intuitions about

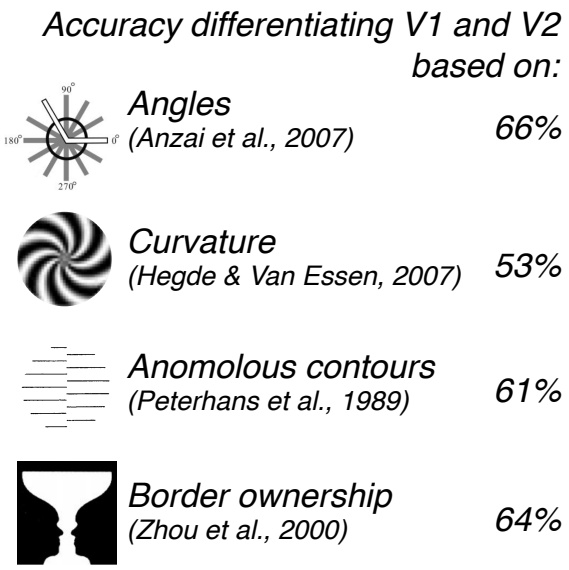

Figure 1.7: Attempts to differentiate V1 and V2 based on novel stimuli. Accuracy was determined by ROC analysis on a single response metric. Data adapted from [106, 91, 169, 259].

what is important in the visual world seem to match what is represented by single neurons. IT neurons are selective for complex objects or visual categories [78, 104, 113, 118, 252, 228]. This highly complex selectivity certainly differentiates their responses from those in early vision, and these neurons can even be linked to specific perceptual behaviors [7]. However, we cannot build interpretable models that predict the response of these individual neurons to arbitrary images until we understand midlevel vision more fully.

One approach to early and midlevel vision that circumvents the problem of stimulus selection entirely is to simply present large numbers of natural images to single neurons [53]. Natural images contain the full complexity of the natural world, so the hard problem of stimulus selection is transformed into the problem of model selection. In areas such as V1 and MT, where there already existed a previous understanding and modeling framework, this approach has been relatively successful [54, 37, 156, 38]. Such studies can often generate good predictions for neural responses to natural images, have confirmed or elaborated observations made with artficial stimuli (such as the importance of suppressive influences), but have mostly not uncovered unknown features of neural representation.

This framework has been less successful in V2. Since the primary input into V2 is from V1, any modeling effort should start with a simulated V1 front end (Fig. 1.8). Willmore and colleagues fit the responses of V2 neurons as linear combinations of wavelets tuned to different orientations and spatial frequencies [250]. The fitted parameters for V2 neurons compared with those fitted to V1 responses did reveal a marked increase in the strength of suppressive tuning. The Willmore model provided adequate fits to V2, but ultimately did not elucidate what specific image properties V2 neurons represent beyond those already represented in

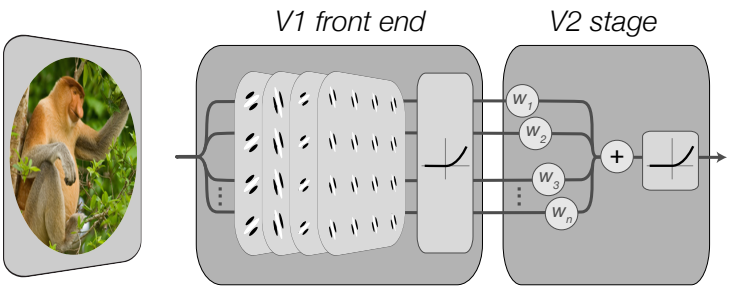

Figure 1.8: Hierarchical modeling of V2 neurons. Responses of a single V2 neuron are predicted from a linear combination of V1-like filter responses. Based on model architectures from [250, 238, 239]

V1. This has generally been the case when fitting similar models to more artificial stimuli [238, 239], as well as when fitting simpler models that operate by taking linear combinations of the pixels of natural images [129]. Thus, although this approach benefits from fitting models that can be applied generally to arbitrary stimuli, they have mostly not differentiated V2 from its V1 inputs and have little hope of providing a link to perception.

A new way of tackling problems of mid- and high-level vision has done away with both choosing stimuli and hand-configuring models. The unprecedented success in recent years of deep convolutional neural networks in object recognition has driven some to suggest that they may serve as a general model of the biological ventral visual stream [254, 28, 112, 253]. A weighted combination of unit outputs at intermediate and higher stages of such a network recently provided the first workable functional model of visual neurons deep in the ventral stream (V4 and IT, [254]). Intriguingly, the better the network was trained to categorize objects, the better it predicted neural responses [254]. This approach has yet to be applied to lower levels of the ventral stream, but may have similar predictive success in V2. However, such a result would ultimately substitute one thing we don't fully understand for another. The performance of these networks has increased rapidly over just the last handful of years, and understanding their internal workings is

still an active area of research [216, 258, 150, 255, 208, 73, 67].

These early results using deep convolutional neural networks suggest that quantitatively accurate models can be built for neurons in higher hierarchical levels of the ventral stream without an understanding of representations in lower levels [254, 112]. Further, some have suggested that intermediate levels of representation may be so complex as to preclude descriptions beyond assigning them a level within a performance-optimized deep network [253]. However, simply finding the right subspace can sometimes obscure understanding of the computations performed by visual neurons. For example, using spike-triggered covariance, one recovers subunits that provide a good prediction for the selectivity of V1 complex cells [187]. However, the subunits themselves often contrain structure that is difficult to interpret. By constructing models based on what we know to be represented in lower levels (here, simple cells), we can build much more efficient, parsimonious, and satisfactory accounts of complex cell responses [240, 241]. Using this bottom-up approach to understanding representations, combined with performance-optimized networks and guided by perception may provide fundamental insights into vision in the coming years [116, 92]. Along these lines, the approach summarized below represents a first step along this path, albeit utilizing a much shallower network.

We posit that to understand V2, we need an approach that is not driven simply by our intuitions about what may be useful at intermediate stages in the ventral visual hierarchy, nor by just aiming to best predict neuronal responses to arbitrary natural images. Instead, we propose a multidisciplinary attack on the problem of midlevel representation, using computational models of lower level representations, insights gleaned from studying the statistics of natural images, and a dynamic interplay between physiology and perception.

1.4 The current approach

In order to approach midlevel vision we have to address the structure of the natural world these neurons must represent. It is well established that sensory neurons are likely to be adapted to efficiently encode the statistics of natural signals [13, 20]. However, the structure of natural images is highly complex and difficult to capture, and there are many ways to establish links between neural processing and environmental statistics [205].

One basic observation is that natural images contain strong spatial correlations. Two neighboring pixels are much more likely to have similar values than two widely separated pixels. We can capture this elementary feature by examining the second-order (covariance) statistics of natural images. If we assume translation invariance across space, the covariance of natural images is captured by the Fourier power spectrum. Natural images have a characteristic relationship between spectral amplitude and spatial frequency, according to the power law $1/f^a$ (with a often between 1 and 2, [223, 182, 205]). V1 neurons capture the spectral properties of natural images well. The receptive fields of V1 neurons are well described by selectivity to a local region of the power spectrum of images [42, 157].

However, natural images are not defined solely by their spectral statistics and contain a great deal of higher-order (beyond second-order) structure. This can been demonstrated by imposing the Fourier power spectrum of a natural image on an image of Gaussian white noise (or by randomizing the phase and maintaining the power spectrum). The resulting image will look radically different and retain none of the recognizable complex features present in natural images [205]. However, a model V1 receptive field convolved with both images will respond the same on

average, because they contain the same amount of all orientations and spatial frequencies (i.e. second-order or spectral statistics).

What higher-order statistics might be needed to capture the perceptually relevant features contained in natural images? Julesz [108] conjectured that when viewing spatially homogeneous texture images, humans would not have the ability to discriminate images that were matched for a sufficient set of statistics [108]. He formulated his theory in terms of the order of pixel statistics, and subsequently found that higher and higher statistical orders were insufficient to match visual perception and abandoned the theory [110, 29, 30, 109]. Working with pixel statistics is challenging as the number of statistics beyond second-order grows exponentially and there is no tractable basis in which to represent them. One way to simplify things and make this approach workable is to limit pixel intensities to just two values (black and white) and generate texture arrays with particular multipoint (beyond second-order) correlations. This creates an image space within which perceptual discrimination can be fully characterized [235]. However, while this statistical description is mathematically rigorous, it cannot be expanded to engage with the full complexity of images in the natural world (but see [222] for one attempt to do so) and is not well aligned to known physiology at the level of V1, prior to areas where such statistics might be represented [256].

The higher-order statistics of natural images can be captured in an alternative framework that is both informed by physiology and can be applied to arbitrary natural images. Specifically, instead of working with the pixels of a natural image we can work with the output of a population of V1-like filters (Fig. 1.9a-c). Examining the outputs of such a mutli-scale and multi-orientation image transformation can yield clues about what structure might be represented at later stages of

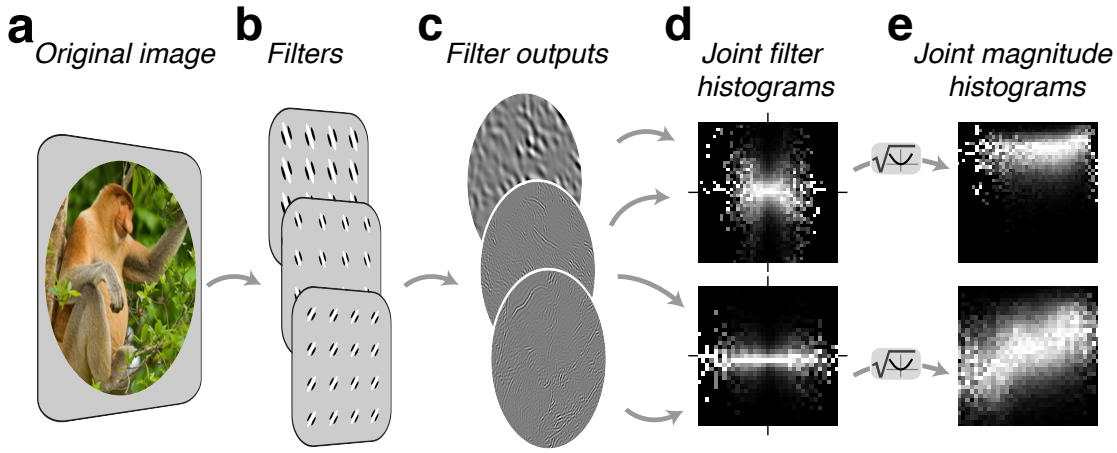

Figure 1.9: Correlations in V1-like filter outputs. A natural image (a) is convolved with 3 filters tuned for orientation and spatial frequency (b). (c) The outputs across the image for three different filters. (d) Examining joint histograms of outputs for filters differing in orientation or scale reveals a lack of correlation but strong dependencies. (e) Plotting the log of the absolute value reveals strong correlations.

the visual hierarchy to capture the attributes of natural images [207]. Heeger and Bergen found that generating synthetic images by matching the marginal statistics of V1-like filter outputs to that of an original recapitulates many features of photographs of natural texture [88]. However, the synthetic images generated through this method also failed to express many key attributes of natural texture images, including strong edges aligned across frequencies, spatial periodicity, and extended contours. These failures suggested that examining dependencies across differently tuned filters might aid in capturing more naturalistic structure.

Although convolving a natural image with V1-like filters decorrelates the second-order pixel statistics, strong dependencies in the output of differently tuned filters remain (Fig. 1.9d, [206]). This is immediately apparent when examining the joint histograms of filters tuned to different orientations (Fig. 1.9d bottom), scales (Fig. 1.9d top), or relative positions. Although their output is not correlated, the joint histogram contains a clear "bow-tie" shape indicating that the response variance of

one filter depends on the activation level of the other. Dependencies across scale, for example, occur in the vicinity of abrupt changes in luminance, and dependencies across orientation and position arise from spatially extended contours [202, 69]. These dependencies can be summarized directly by a correlation if considering the filter magnitudes (absolute values, Fig. 1.9e). Thus, higher-order statistics in natural images can be captured by second-order statistics applied to the output of a V1 front end.

What kind of visual attributes can be encapsulated by correlations in the magnitude of filter outputs? Portilla and Simoncelli found that the appearance of many visual textures can be recreated by imposing a well-chosen set of correlations (or more accurately covariances) across filter outputs (Fig. 1.10, [172]). To analyze an original image, they first decomposed it into the outputs of linear and energy filters tuned to different orientations and scales (Fig. 1.10b [207]). The second stage of the model computes local correlations between the output of energy filters tuned to different orientations and scales, and different relative positions within a subband (Fig. 1.10a). The autocorrelation and cross-scale phase correlation of linear filters are also included, as well as the marginal statistics of the pixels of the original image. Following this analysis stage, They then use a synthesis algorithm to iteratively modify a image of Gaussian white noise until it is exactly matched for this set of statistics [172].

The resulting images (Fig. 1.10d) capture many of the naturalistic features of original texture images (Fig. 1.10c). Depending on the complexity of the originals, humans can be quite poor at discriminating between original and synthetic texture images [16]. However, the model can only recreate the appearance of natural texture images. When applied to inhomogeneous natural images containing

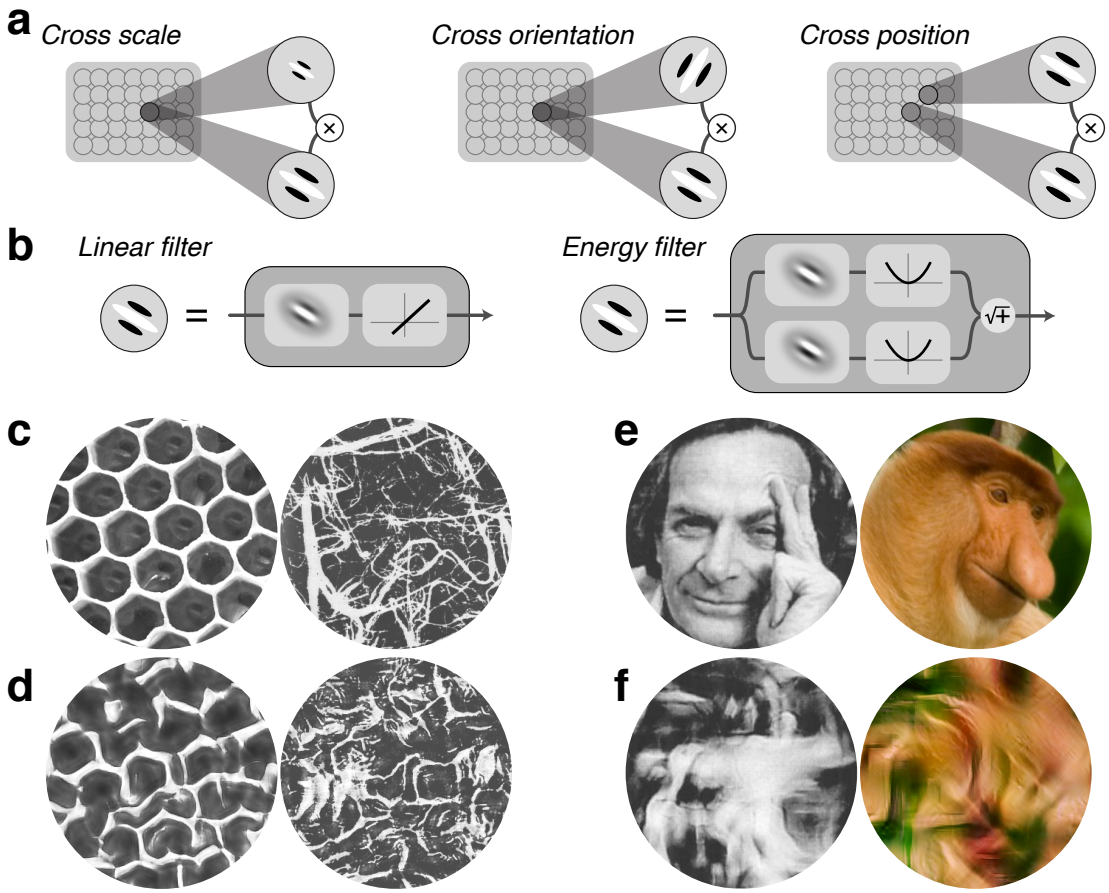

Figure 1.10: Texture analysis and synthesis using the Portilla and Simoncelli model [172]. (a) Cross-scale, cross-position and cross-orientation correlations are computed by taking products of localized V1-like filter responses. Each circle represents an image location. Filters at each location are tuned to orientation and frequency, and compute either linear or energy responses. (b) Linear filters are sensitive to phase, akin to V1 simple cells; energy filters compute the square root of the sum of squared responses of two phase-shifted filters (in quadrature pair) and are thus insensitive to phase, akin to V1 complex cells [4]. For both filter types, products are averaged across spatial locations to yield correlations. (c) Original natural textures and (d) naturalistic images synthesized using the Portilla and Simoncelli algorithm. (e) non-textural natural images and (f) their synthesized versions.

recognizable objects and features (Fig. 1.10e), the model fails to synthesize images that could be confused with the originals (Fig. 1.10f).

This failure comes from the fact that the statistics are measured and imposed

averaged over the entire spatial extent of the image. This averaging destroys the global image structure that defines objects and complex shapes. This might suggest that much more complicated measurements (or higher-order statistics) would be required to match the appearance of natural images. Instead, a simple set of statistics (such as those used by Portilla and Simoncelli) may be sufficient when they are only averaged over relatively small regions of the image. Visual textures are ubiquitous in natural images, and capturing their features locally may be a good candidate for a representation of intermediate complexity from which more complex visual attributes can be extracted. In fact, when viewing images through the visual periphery they tend to take on a more textural character [122]. We lose our ability to recognize objects, faces, and letters when lots of visual structure is "crowded" together in the visual periphery [167, 168]. Many have suggested that a summary statistic representation operates in the periphery to explain such perceptual limitations [122, 166, 167, 76, 17].

Freeman and Simoncelli tested this idea directly by building a model that computed the statistical dependencies in the output of a V1-like representation, but combined it with the fact that visual receptive fields scale with eccentricity (Fig. 1.5). They computed the same statistics from the Portilla dn Simoncelli texture model (Fig. 1.11, [172]) within pooling regions that grow in size from a central point (akin to the model's fovea) in an original natural photograph (Fig. 1.11g). They then adjusted an image of white noise until the statistics within each pooling region were exactly matched to the original (Fig. 1.11b). Varying the scaling of pooling regions creates synthetic images with more or less distortion at a given distance from the image center. Observers who fixated the central point were unable to discriminate between two statistically matched samples when the

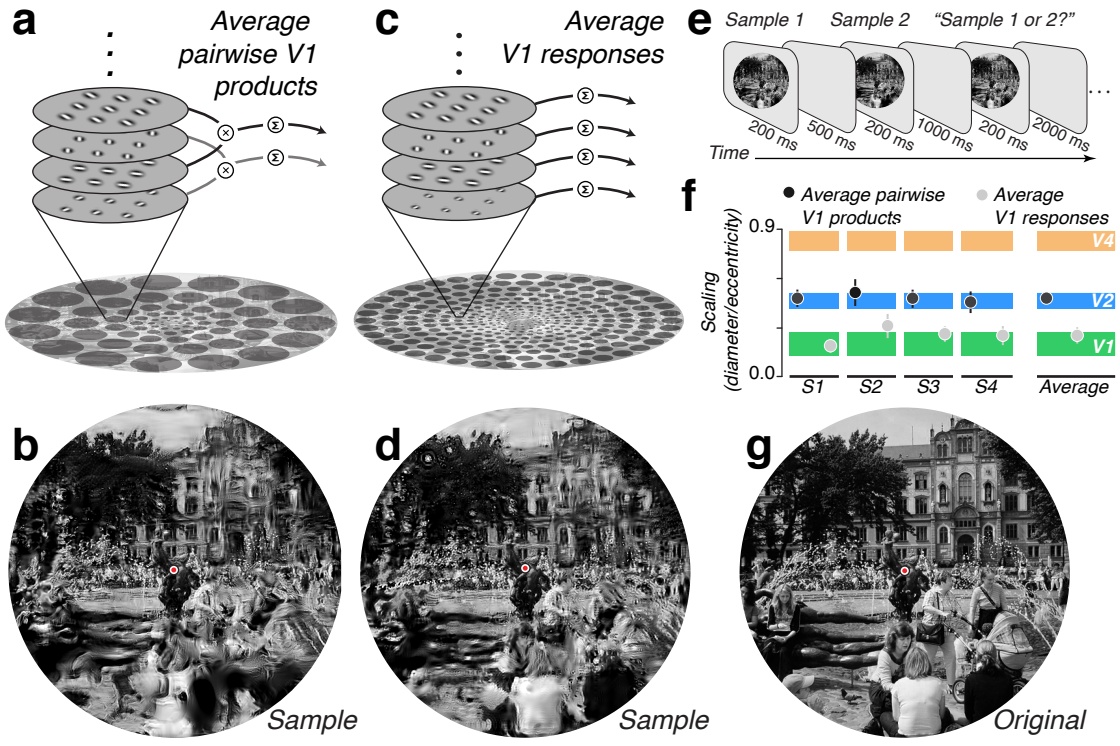

Figure 1.11: Metamers of the ventral stream. (a) higher-order statistical dependencies from an original natural image are captured within each pooling region by computing the average pairwise products of a V1 front end. (b) Synthetic image matched exactly for higher-order statistical dependencies within V2 sized pooling regions. (c) Spectral (second-order) statistics alone are captured within each pooing region by computing the average output of a V1 front end. (d) Synthetic image matched exactly for spectral statistics within V1 sized pooling regions. (e) Psychophysical ABX task. (f) Summary of psychophysical scaling for the two types of images shown in (b) and (d). Colored horizontal bars represent 95% confidence interval on receptive field scaling as measured physiologically in each visual area. (g) Original image. Adapted from [60]

pooling regions scaled with eccentricity at a rate that matched the receptive field sizes of V2 neurons (Fig. 1.11ef). Importantly, they confirmed that the joint statistics of differently tuned V1 filters were specifically represented in V2 sized pooling regions. Observers were able to discriminate different samples generated with a simpler model containing only the average output of different V1-like filters at much smaller pooling region sizes (Fig. 1.11c-f). But again, they became unable

to discriminate images matched for this V1-like representation when the pooling regions matched the size of V1 receptive fields.

All these previous results suggest that the higher-order statistics underlying the appearance of naturalistic textures provide a potentially effective foundation for investigating midlevel form vision and its neural basis. The biologically inspired parameterization of these features [172], as well as psychophysical experiments [60] point to V2 as a potential neural locus for the representation of these features. What follows in this dissertation is a series of physiological investigations to test this hypothesis. Our results suggest a new way of approaching midlevel sensory computation, not just in area V2 but further downstream in the ventral hierarchy, and not just in vision but in other sensory domains as well.

Chapter 2

A functional and perceptual signature of V2

2.1 Introduction

The perception of complex visual patterns emerges from neuronal activity in a cascade of areas in the primate cerebral cortex. Neurons in the primary visual cortex (V1) represent information about basic image features like local orientation and spatial scale. Downstream areas contain neurons sensitive to more complex properties, especially those found in behaviorally relevant, natural images. But sensitivity to these naturalistic structures requires transformations of basic visual signals, which have been difficult to characterize in computational or physiological terms.

The role of the second visual area (V2) has been particularly enigmatic. V2 is the largest extrastriate visual cortical area in primates, and its responses depend on feedforward input from V1 [190, 210]. Neurons in V2 presumably combine and elaborate signals from V1 to encode image features that V1 does not, but the responses of V2 neurons to most "artificial" stimuli, including gratings, angles,

curves, anomalous contours, and second-order patterns are largely similar to the responses of neurons in V1 [169, 89, 91, 120, 132, 106, 11, 196]. Reliable responses to border ownership and relative disparity are more prevalent in V2 than in V1 [259, 220], and V2 neurons exhibit stronger tuned suppression [250], but neither of these properties reliably and robustly distinguish V1 and V2 neurons.

We wondered whether the responses of V2 cells encode aspects of natural image structure. A ubiquitous property of natural images is that they contain orderly structures which create strong statistical dependencies across the outputs of filters — similar to the responses of V1 cells — tuned to different positions, orientations, and spatial scales [206, 193, 111]. Dependencies across scale, for example, occur in the vicinity of abrupt changes in luminance, and dependencies across orientation and position arise from spatially extended contours [202, 69]. The character and extent of these dependencies varies for different classes of images. Many artificial stimuli lack them; in photographs of scenes and objects, they are present but sparse, non-uniform, and difficult to control. But a sub-class of natural images — visual textures — contain these dependencies in a form that can be captured parametrically [172], and previous psychophysical investigations suggest that V2 may be the locus for representing them [60].

We found a novel and distinctive property of V2 cells by measuring their responses to controlled naturalistic texture stimuli. We first captured the statistical dependencies in natural texture photographs by computing correlations among the outputs of V1-like filters tuned to different positions, orientations, and spatial scales. We then generated families of homogenous textures containing the same statistical dependencies. There was a unique response to these texture stimuli in V2 but not V1, in both macaque and human, that reliably predicted perceptual

sensitivity to the same stimuli. A large-scale crowdsourced psychophysics experiment revealed the statistical dependencies most relevant for perception and — by extension — selective responses in V2. Together, these findings situate V2 along a cascade of computations that lead to the representation of naturally occurring patterns and objects.

2.2 Methods

2.2.1 Model and synthesis of stimuli

Model

Here we describe aspects of the model and stimulus generation common to all experiments. Further details of stimulus presentation for each experimental method are presented separately, below. Stimuli in all experiments were generated using the texture analysis-synthesis procedure described in [172] (code and examples available at `http://www.cns.nyu.edu/~lcv/texture/`) . We began with an ensemble of diverse natural homogenous black and white photographs of visual textures, drawn from both commercial and personal databases. Each original texture photograph served as the basis for a "texture family." Most of our experiments used 15 texture families, selected to vary in the extent to which each texture differed from an image of spectrally matched noise. The crowdsourced psychophysical experiments (Fig. 2.20, 2.21, 2.22) used an additional 479 texture families, selected only to avoid duplicates and images with entirely blank regions (e.g. clouds). For each texture family, we computed the model parameters on the original photograph, by processing the image with a multi-scale multi-oriented bank of filters with 4 orientations and 4 spatial scales. For each filter, we computed the real, lin-

ear output and the energy (square root of sum of squared quadrature pair outputs), analogous to the responses of simple and complex V1 cells. We then computed pairwise products across filter responses at different positions (within each orientation and scale), across different orientations, and across different scales. All of these pairwise products were averaged across the spatial extent of the image, yielding correlations. We additionally computed spectral statistics (average energy within each band of the pyramid) and marginal pixel statistics (skew and kurtosis).

Synthesis

After computing the model responses on an original image, we synthesized 15 new samples by initializing 15 different images of Gaussian white noise, and adjusting each using gradient descent (specifically, gradient projection) until it matched the model parameters computed on the original image. Because the dimensionality of the image was larger than the number of parameters, this process yielded multiple random high-entropy samples that were statistically identical in terms of the model parameters. Convergence of all parameter groups was monitored and ensured, and the number of synthesis iterations used (50) was far more than typically required. For each texture family, we also generated spectrally matched "noise" images by randomizing the phase but matching the complete two-dimensional power spectra. This procedure yielded nearly identical results to iteratively matching the power averaged within each band of the multi-scale multi-oriented filter bank. We performed noise synthesis separately on each naturalistic texture sample to generate 15 samples. For psychophysical experiments, we generated stimuli that spanned a "naturalness" axis between noise (0) and naturalistic (1). For each texture family we computed the model parameters $\vec{P}_{nat}$ on the original natural photograph and pa-

rameters $\vec{P}_{noise}$ on a spectrally matched noise image, and then linearly interpolated the model parameters between the two endpoints, $\vec{P}_{interp} = \Delta\vec{P}_{nat} + (1-\Delta)\vec{P}_{noise}$. For each interpolated parameter vector, we used the same synthesis procedure to generate 15 samples. Pilot experiments suggested that the distribution of thresholds across texture families was approximately normally distributed in the log domain, so we sampled the naturalness axis with 10 values of equally spaced on a logarithmic scale. For laboratory psychophysics, we used a range of 0.04-0.8; for crowdsourced psychophysics, we used a range of 0.1-1.0 (see below).

2.2.2 Physiology

Recording

We recorded from 13 anesthetized, paralyzed, adult macaque monkeys (2 M. Nemestrina and 11 M. Cynomolgus). Our standard methods for surgical preparation have been documented in detail previously [34]. We maintained anesthesia with infusion of sufentanil citrate (6-30 $\mu g/kg/hr$) and paralysis with infusion of vecuronium bromide (Norcuron; 0.1 $mg/kg/hr$) in isotonic dextrose-Normosol solution. We monitored vital signs (heart rate, lung pressure, EEG, body temperature, urine volume and specific gravity, and and end-tidal pCO_2) and maintained them within the appropriate physiological range. The eyes were protected with gas permeable contact lenses and refracted with supplementary lenses chosen through direct ophthalmoscopy. At the conclusion of data collection, the animal was killed with an overdose of sodium pentobarbital. All experimental procedures were conducted in compliance with NIH Guide for the Care and Use of Laboratory Animals and with the approval of the New York University Animal Welfare Committee. We made a craniotomy and durotomy centered approximately 2-4mm

posterior to the lunate sulcus and 10-16mm lateral and individually advanced several quartz-platinum-tungsten microelectrodes (Thomas Recording) into the brain at an angle 20 degrees from vertical. We distinguished V2 from V1 on the basis of depth from the cortical surface and changes in the receptive field location of recorded units. In an effort to obtain an unbiased sample of single units, we made extracellular recordings in V1 and V2 from every single unit with a spike waveform that rose sufficiently above noise to be isolated, and we fully characterized every unit that demonstrated a measurable visually evoked response to gratings or naturalistic texture. Data are reported from every unit for which we completed characterization (see below). The receptive fields of most units were between 2 and 5 degrees eccentricity, but our estimates of eccentricity were not sufficiently precise to include in analyses.

Visual stimulation

We presented visual stimuli on a gamma-corrected CRT monitor (Eizo T966; mean luminance, 33 cd/m^2) at a resolution of 1280×960 with a refresh rate of 120Hz. Stimuli were presented using Expo software on an Apple Macintosh computer. For each isolated unit, we first determined its ocular dominance and occluded the non-preferred eye. We used drifting sinusoidal gratings to characterize the basic receptive field properties of each unit, including tuning for orientation and direction, spatial and temporal frequency, size, and contrast. We then presented the texture stimuli. We used a set of 15 texture families, and generated 15 samples for each texture family for a total of 225 images. 15 spectrally matched noise samples of the 15 families were also presented. The 450 unique images making up our stimulus ensemble were presented in pseudo-random order for 100 ms each,

separated by 100 ms of mean luminance. Each image was presented 20 times. Images were presented to every unit at the same scale and at a size of 4 degrees within a raised cosine aperture. We chose a 4 degree aperture to be larger than all the receptive fields at the eccentricities from which we typically record. Nearly all recorded units had receptive fields smaller than 4 degrees, and the majority were less than 2 degrees. For a subset of V1 and V2 neurons we additionally presented stimuli in a smaller aperture matched to the receptive field size of that unit. The aperture diameter was set to be the grating summation field as measured with full contrast drifting gratings [34]. We ran the full texture stimulus ensemble within this aperture although typically with only 5-10 repeats per image.

Analysis

The full stimulus ensemble consisted of 450 images presented 20 times each. All analyses were performed after averaging spiking responses across those 20 repeats, and also averaging responses across the 15 samples. Depending on the analysis, responses were further averaged across texture family, neurons, and/or a temporal window. Response time courses were computed by counting spikes within a sliding, non-overlapping 10 ms window. Time courses were always averaged across texture families (Fig. 2.4ab). For the population average plot (Fig. 2.4b), time courses for each neuron were first normalized by dividing by each neuron's maximum response across all texture families and time points, but after averaging responses evoked by the 20 repeats of each of the 15 images within the same texture family. A modulation index was computed as the difference in firing rate between naturalistic and noise divided by the sum. The index was computed separately for each texture family. For time course plots (Fig. 2.4c), modulation was computed within 10

ms windows. In all other cases, firing rates were first averaged within an 100 ms window following response onset, and the modulation index was computed on those rates. Response onset was determined by inspection as the time point eliciting a response above baseline; results were nearly identical when using a quantitative criterion based on the standard deviation of the response. Finally, modulation indices were additionally averaged across neurons (Fig.2.5b) or across texture families (Fig.2.7).

Basic receptive field properties for each neuron — e.g., receptive field size, surround suppression — were determined offline by using maximum likelihood estimation to fit an appropriate parametric form to each tuning function. These fits were only obtainable for a subset of neurons (81% in V1, 74% in V2) due to incomplete characterization arising from time constraints during the experiment.

2.2.3 fMRI

Subjects

Data were acquired from three healthy subjects with normal or corrected-to-normal vision (all male; age range, 26-30 years). Two subjects were authors. Experiments were conducted with the written consent of each subject and in accordance with the safety guidelines for fMRI research, as approved by the University Committee on Activities Involving Human Subjects at New York University. Each subject participated in at least three scanning sessions: one session to obtain a set of high-resolution anatomical volumes, one session for standard retinotopic mapping (single wedge angular position, and expanding ring eccentricity), and one session to measure differential responses to naturalistic and spectrally matched noise stimuli. Two subjects participated in two additional scanning sessions, one using tex-

ture families derived from the crowdsourced psychophysical experiment (described below), and one to measure differential responses between natural photographic images, naturalistic textures, and spectrally matched noise images.

Stimuli

Stimuli were presented using Matlab (MathWorks) and MGL (available at `http://www.justingardner.net/mgl`) on a Apple Macintosh computer. Stimuli were displayed via an LCD projector (Eiki LC-XG250) onto a back-projection screen in the bore of the magnet. Subjects laid supine and viewed the stimuli through an angled mirror. All images were presented within a suitably vignetted annular region (inner radius, 2 degrees; outer radius, 8 degrees). We used textures that approximately matched in scale the presentation conditions in the electrophysiological and psychophysical experiments.

Protocol

Blocks of naturalistic and spectrally matched noise stimuli were presented in alternation. Within each 9 s block, a random sequence of images from one texture family were presented at 5 Hz. Each run consisted of 20 blocks: 10 naturalistic, 10 noise. Different texture families were presented in separate runs, and multiple runs were performed within each session. Subjects performed two runs for each texture family. In each session, a separate localizer run was used to define retinotopic subregions corresponding to the stimulus region. Within each 9 s block of the localizer run, a random sequence of both naturalistic and noise stimuli were presented within the stimulus annulus or the region complementary to the annulus. Each run consisted of 40 blocks: 20 annulus, 20 anti-annulus.

Task

Observers performed a demanding two-back detection task continuously throughout each run to maintain a consistent behavioral state, encourage fixation, and divert attention from the peripheral stimulus. Without attentional control, we have reported large and variable attentional signals in visual cortex [177]. Digits (0 to 9) were displayed continuously at fixation, changing every 400 ms. The observer used a button press to indicate whether the current digit matched the digit from two steps before.

Preprocessing

All preprocessing and analyses were implemented in Matlab using mrTools (`http://www.cns.nyu.edu/heegerlab/?page=software`). The anatomical volume acquired in each scanning session was aligned to the high-resolution anatomical volume of the same subject's brain, using a robust image registration algorithm [147]. Data from the first half cycle (eight frames) of each functional run were discarded to minimize the effect of transient magnetic saturation and allow the hemodynamic response to reach steady state. Head movement within and across scans was compensated for using standard procedures [147] The time series from each voxel was high-pass filtered (cutoff, 0.01 Hz) to remove low-frequency noise and drift [211].

Analysis

We performed two complementary analyses of fMRI responses to alternating blocks of naturalistic texture and noise stimuli, one to visualize responses, and a second to quantify them for statistical analyses (and for comparisons to psychophysics and physiology). First, for each voxel, response time courses were averaged across tex-

ture families, and fit with a sinusoid with period matched to the block alternation (9 s). The coherence between the best-fitting sinusoid and the average time series was used to assess the statistical reliability of differences in cortical activity evoked by naturalistic and noise stimuli, visualized on flattened maps of the occipital lobe (Fig. 2.12, 2.14a, 2.16).

To quantify responses, we computed an fMRI modulation index, analogous to the index used for single unit measurements. We computed the index as the ratio of two response amplitudes: the amplitude of differential responses to naturalistic versus noise (texture minus noise), and the amplitude of differential responses to naturalistic and noise together versus a blank screen (texture plus noise). To obtain the numerator (texture minus noise), for each texture family, we averaged the time course of each voxel across repeated runs, and then projected it onto a unit-norm sinusoid having period matched to the stimulus alternation, and phase given by the responses to the localizer scan (see above). The reference phase provided an estimate of the hemodynamic delay, and was computed separately for each visual area. The amplitude of projection isolated the component of the response time course that responded positively and differentially to the naturalistic texture stimuli [86]. To obtain the denominator (texture plus noise), we projected the response time courses from the localizer scan onto a unit-norm sinusoid with the same reference phase. The amplitude of this projection captured the combined response to texture and noise images together, because the localizer scan presented both (randomly interleaved) alternating with a blank screen. Both response amplitudes (texture minus noise, and texture plus noise) were averaged across voxels, and their ratio yielded a modulation index for each visual area. fMRI modulation indices were then either averaged across texture families (Fig. 2.13) or analyzed

separately (Fig. 2.14b, 2.19b, 2.21b). Results were qualitatively similar (and supported the same conclusions) when this fMRI modulation index was replaced by either coherence, or the texture minus noise response amplitudes (without division by texture plus noise response amplitudes).

MRI acquisition

MRI data were acquired on a Siemens 3T Allegra head-only scanner using a head coil (NM-011; Nova Medical) for transmitting and an eight-channel phased array surface coil (NMSC-071; Nova Medical) for receiving. Functional scans were acquired with gradient recalled echo-planar imaging to measure blood oxygen level dependent changes in image intensity [158]. Functional imaging was conducted with 24 slices oriented perpendicular to the calcarine sulcus and positioned with the most posterior slice at the occipital pole (1500 ms repetition time; 30 ms echo time; 72^{o} flip angle; $2 \times 2 \times 2$ mm voxel size; 104×80 voxel grid). A T1-weighted magnetization-prepared rapid gradient echo anatomical volume (MPRAGE) was acquired in each scanning session with the same slice prescriptions as the functional images (1530 ms repetition time; 3.8 ms echo time; 8^{o} flip angle; $1 \times 1 \times 2.5$ mm voxel size; 256×160 voxel grid). A high-resolution anatomical volume, acquired in a separate session, was the average of three MPRAGE scans that were aligned and averaged (2500 ms repetition time; 3.93 ms echo time; 8^{o} flip angle; $1 \times 1 \times 1$ mm voxel size; 256×256 voxel grid). This high-resolution anatomical scan was used both for registration across scanning sessions and for gray matter segmentation and cortical flattening.

Defining retinotopic regions of interest

Each subject participated in a standard retinotopic mapping experiment, described in detail previously [119, 63]. The data were analyzed, following standard procedures to identify meridian representations corresponding to the borders between retinotopically organized visual areas V1, V2, V3, and V4. There is some controversy over the exact definition of human V4 and the area just anterior to it; we adopted the conventions proposed by [246]. We used data from an independent localizer scan (see above) to further restrict each visual area to only those voxels responding to the stimulus annulus with coherence of at least 0.25. Qualitatively similar results were obtained using higher or lower thresholds.

2.2.4 Psychophysics (Laboratory)

Observers

Three observers with normal or corrected-to-normal vision participated in the experiments (all male; age range, 26-30 years). Protocols for selection of observers and experimental procedures were approved by the Human Subjects Committee of New York University. Two observers were authors. The other was naive to the purpose of the experiment.

Stimuli

Stimuli were presented on a 41×30cm flat screen CRT monitor at a distance of 46cm. Texture images were presented within vignetted 4 degree circular patches at three locations equidistant from fixation, each 4 degree eccentricity (one above fixation, one to the lower left, and one to the lower right). A 0.25 degree fixation

square was shown throughout the experiment.

Task

Every trial of the 3AFC "oddity" task presented three different image within the three patches: two images were spectrally matched noise and one was naturalistic, or one was noise and two were naturalistic. The "naturalness" of the naturalistic texture(s) varied across trials, spanning ten points between 0.4 and 0.8, equally spaced on a log scale. If two naturalistic textures were presented on a trial, they had the same level of naturalness. Image patches were presented for 600 ms, after which observers had 1 sec to indicate with a keypress which of the three was the odd one out. There was no feedback during the experiment. Before the experiment, each observer performed a small number of practice trials ($\sim$10) with feedback to become familiar with the task. Different texture families were run in separate blocks. Each observer performed 480 trials in each block; the order of conditions and location of the target were appropriately randomized and counterbalanced. Blocks were performed in random order for each subject. Data were collected from 15 texture families.

Analysis

For each texture family, we fit the parameters of a Weibull function that maximized the likelihood of the psychometric data. The function was parameterized with a threshold, slope, and lapse rate [248]. Estimated lapse rates were typically very small (mean 0.01, maximum 0.06). Threshold was converted to its reciprocal (sensitivity) for all subsequent analyses, and statistics, e.g. correlations, were computed in the log domain (Fig. 2.19, 2.21, 2.22).

2.2.5 Psychophysics (crowdsourced)

Observers

Several hundred observers ("Turkers") were recruited for experiments through Amazon.com's Mechanical Turk website. Each was paid $0.40 for approximately 5 minutes of their time. Payment was made so long as Turkers completed the task, regardless of performance. Demographic data were not collected, but demographic studies of the Mechanical Turk [164] suggest that our sample reflected gender and age diversity. Participation was restricted to those Turkers achieving 95% approval rating on other Mechanical Turk tasks. Protocols for selection of Turkers and experimental procedures were approved by the human subjects committee of New York University. All Turkers signed an electronic consent form at the beginning of the experiment. We ensured that 10 unique Turkers completed the task for each texture family, but we did not prevent the same Turkers from completing the task for multiple texture families.

Stimuli

We developed a version of our 3AFC task for display in a web browser (see example at `http://www.jeremyfreeman.net/public/turk/code/?csv=tex-018-files.csv`), using Javascript and HTML. Each trial began with 700 ms blank period, followed by a 600 ms stimulus presentation, and a second 700 ms blank period. As in the laboratory version of the experiment, images were presented in three patches equidistant from fixation. A small red fixation dot was shown throughout the experiment. After the second blank, three arrows were presented near fixation pointing towards the three possible target locations. Turkers were instructed that

"One image will look different from the other two — your task is to identify it by clicking the black arrow that points to it." There was no other explanation of the nature of the stimuli or the conditions. We were unable to verify or control viewing distance, size, eccentricity, or presentation time. However, data obtained from the crowd and from the lab were comparable (Fig. 2.20b), suggesting that such variations were unimportant, at least with respect to this stimulus and task.

Task

Trial types for the 3AFC task were similar to those in the laboratory experiment, except naturalness was varied across ten points equally spaced on a logarithmic scale between between 0.1 and 1.0, instead of 0.04-0.8. This range was chosen because pilot experiments suggested moderately higher thresholds compared to the laboratory data. Each Turker performed 60 trials, and different texture families were run separately. There was no feedback during the experiment, but Turkers performed 6 trials at the beginning with 1.0 naturalness, and were told that these initial trials would be easier than the others. Data were collected from 494 texture families.

Analysis

Each Turker and texture family yielded a psychometric function, based on six trials for each of ten levels of naturalness. Typically, for each texture family, a small number of Turkers performed at or near chance at all naturalness levels, suggesting that they may not have been performing the task appropriately. If data from all Turkers were averaged, the influence of these Turkers would have yielded fitted psychometric functions with unreasonably high lapse rates [248]. As an alterna-

tive, we described the data using a mixture model, with one common psychometric function and an individual lapse rate for each Turker, based on an approach developed in a related problem setting [41]. The analysis inferred the quality of individual Turkers and appropriately weighted their contribution to estimates of threshold. Although we consider this approach appropriate for these data, simple averaging of Turker responses yielded qualitatively similar results.

Validation of perceptual-neuronal relationship

We used the crowdsourced sensitivities to validate the relationship between perceptual sensitivity and neuronal response as measured both with fMRI and in single units. From the distribution of 494 sensitivities, we selected 20 texture families that sampled the range of sensitivity, emphasizing the extremes (Fig. 2.20c), and not including the 15 used previously. In two human subjects, we performed an additional fMRI experiment measuring responses to these 20 texture families. In one monkey, we recorded responses from 16 single units in V2 and 11 single units in V1 to 17 of the texture families (3 of the families were excluded due to experimental time constraints). Experimental procedures and analyses for both fMRI and single-unit experiments were otherwise identical to those described above.

2.2.6 Predicting perceptual sensitivity from texture statistics

All naturalistic textures were generated by matching an image for a particular set of higher-order image statistical parameters derived from an original texture photograph. We used a combination of principal components analysis and multiple linear regression to relate diversity in these parameters to diversity in perceptual sensitivity — and, by extension, neuronal response in V2. We began by comput-

ing all parameters for each texture family. The parameters were appropriately transformed so that all varied linearly in image contrast; for example, by taking the signed square root of correlations. Parameters were then Z-scored so that, for each parameter, the mean of its value across the images was 0, and the standard deviation was 1. We then grouped the parameters as follows: (1) marginal statistics (skew and kurtosis) (2) spectral statistics (average energy in each subband), (3) correlations of linear filter responses at neighboring locations, (4) correlations of linear filter responses at neighboring scales, (5) correlations of energy filter responses at neighboring orientations, (6) correlations of energy filter responses at neighboring locations, (7) correlations of energy filter responses at neighboring scales. For each group of parameters g, we constructed the $494 \times p_g$ matrix P_g containing the p_g parameters in that group for the 494 texture categories. We then reduced the dimensionality of each group of parameters separately using principal components analysis, projecting each parameter matrix into the space spanned by the first k components, yielding a $494 \times k_g$ matrix $\hat{P}_g$. We used the k components required to capture 70% of the variance in each parameter group (typically between 2 and 6, at most 10), for a total of 35 components across all groups. Overall predictive performance was similar when using only 1 component per parameter group, but that would have made it inappropriate to compare the predictive power of the different parameter groups (see below).

Having reduced the dimensionality of each parameter group, we obtained a combined predictor matrix X, with 494 rows and 35 columns, and used multiple linear regression to predict sensitivity to the parameters. We added a column of ones to the matrix (to account for a constant offset), and solved for the weights $\hat{\vec{b}}$ that minimized the squared error,

$$E = ||X\vec{b} - \vec{y}||^2$$

where $\vec{y}$ is a vector of log sensitivities for each of the texture families (as mentioned above, we worked in the log domain because log sensitivities were approximately normally distributed). We removed from analysis any families where thresholds were estimated as greater than 1.0 or less than 0.0 naturalness (only 4% of families), to avoid the influence of outliers arising from unstable threshold estimates.

R^2 for the linear model was used to assess prediction accuracy. R^2 was computed for the full model fit, as well as using 10-fold cross-validated, where the model was fit to 9/10 of the data and R^2 was evaluated on the remaining 1/10, and R^2 was averaged over different splits.

Three complementary procedures were used to assess the relative importance of the different parameter groups in predicting sensitivity. When parameter groups are correlated, as ours were, there is no objective decomposition of R^2, but for our primary analysis we used a well-established procedure known as "averaging over orderings" [77]. For each parameter group, a difference in R^2 is computed for two models, only one of which contains the group. This differential R^2 depends on the order in which the different parameter groups were added to the model, as well as the size of the model when the group was added, so its value is averaged over all possible order permutations and model sizes. The resulting estimates of R^2 for each parameter group exactly partition the full model's R^2 (Fig. 2.22).

As a complementary analysis, we assessed the marginal predictive accuracy

of each parameter group by computing R^2 when including each parameter group on its own. We also assessed the conditional predictive accuracy of each parameter group by computing the difference in R^2 for two models containing all parameter groups, with or without the group of interest. Both additional analyses yielded qualitatively similar results to the averaging-over-orderings procedure, in particular, emphasizing the importance of cross-scale dependencies of energy filter responses.

2.3 Generating naturalistic texture stimuli

For each of several original photographs of visual texture, we transformed samples of Gaussian noise to synthesize new images with the statistical properties of

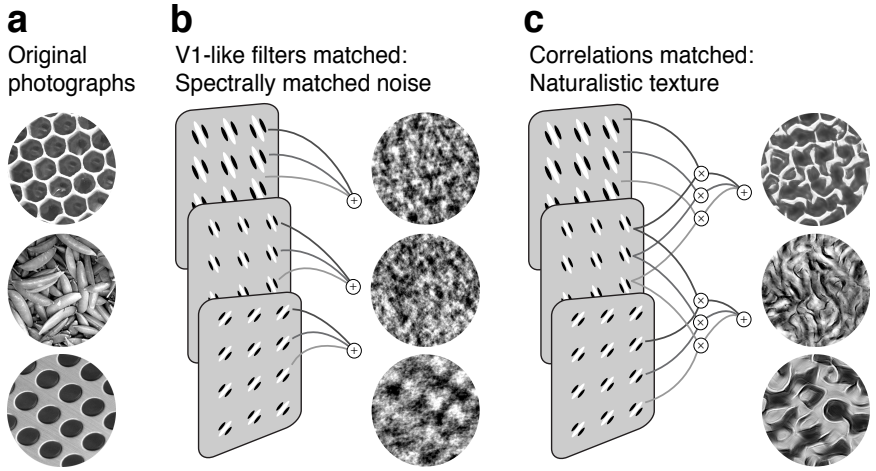

Figure 2.1: Analysis and synthesis of naturalistic textures. (a) Original texture photographs. (b) Spectrally matched noise images. The original texture is analyzed with linear filters and energy filters (akin to V1 simple and complex cells, respectively) tuned to different orientations, spatial frequencies, and spatial positions. Noise images contain the same spatially averaged orientation and frequency structure of the original. (c) Naturalistic texture images. Correlations are computed by taking products of linear and energy filter responses across different orientations, spatial frequencies, and positions. Images are synthesized to match both the spatially averaged filter responses and the spatially averaged correlations between filter responses.

the original [88, 172] (Fig. 2.1, 2.3). For each original texture, we generated two sets of stimuli using different statistics: "spectrally matched noise images" and "naturalistic texture images". Spectrally matched noise images were synthesized using phase randomization, i.e., by computing the Fourier transform, randomizing the phase values, and then inverting the Fourier transform. This is approximately equivalent to measuring and matching the spatially averaged responses of linear and energy filters (akin to V1 simple and complex cells, respectively) selective for different orientations, positions, and spatial scales. The resulting synthetic images had the same overall orientation and spatial-frequency content as the original (i.e., the same spectral properties) but lacked its higher-order statistical dependencies (Fig. 2.1a,b). Naturalistic texture images were generated by additionally matching correlations between filter responses (and their energies) across orientations, positions, and spatial scales (Fig. 2.1a,c). We used an iterative procedure (Fig. 2.2) to match the spatially averaged filter responses, the correlations between filter responses, and also the mean, variance, skewness, and kurtosis of the pixel luminance

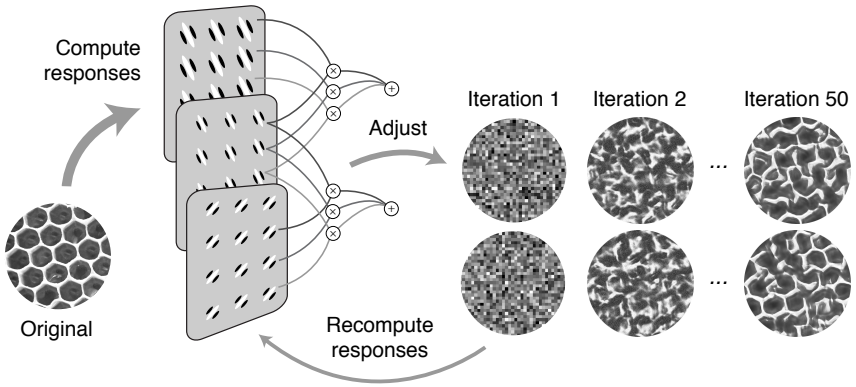

Figure 2.2: Synthesis of naturalistic textures begins with Gaussian white noise, and the noise is iteratively adjusted using gradient descent until analysis of the synthetic image matches analysis of the original (see [172]). Initializing with different samples of Gaussian noise yields distinct but statistically similar images.

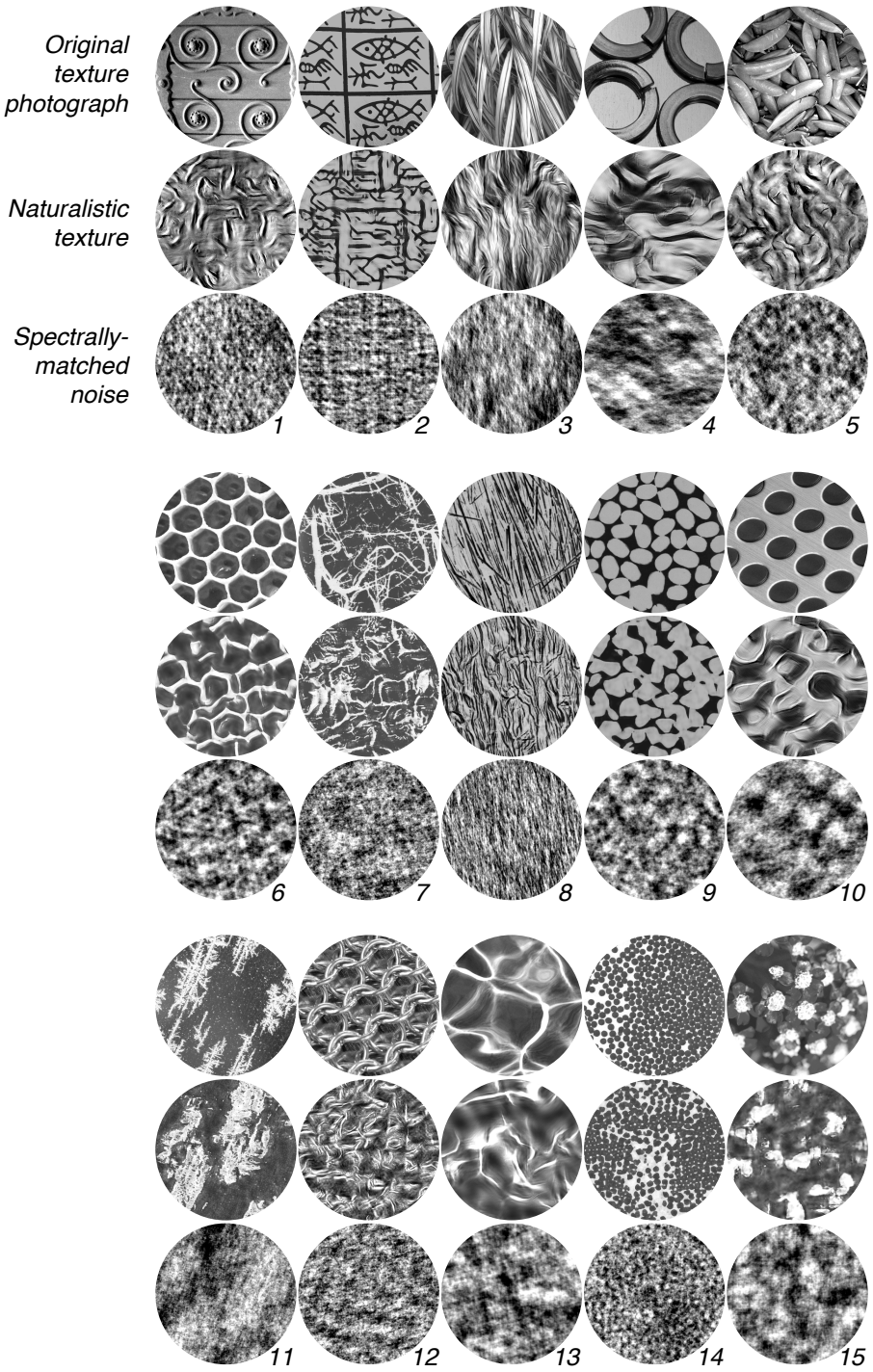

Figure 2.3: Original texture photographs, synthetic naturalistic textures, and spectrally matched noise images, for the 15 texture families used in our primary experiments. Families are sorted by the differential neuronal response they evoked in V2 (as in 2.5)

distribution ("marginal statistics"). Synthetic images matched for these properties contain many complex naturalistic structures seen in the original photograph [172], readily recognizable by human observers [16].

We synthesized images based on 15 original texture photographs, yielding 15 different "texture families"; for each original, we made ensembles of self-similar naturalistic texture samples, each different in detail but all having identical statistical dependencies and containing similar visual properties (Fig. 2.3). Since each of these 15 texture families was based on a different original photograph, they varied in their appearance, and in the form and extent of their higher-order statistical dependencies.

2.4 Differentiating V2 from V1 in macaque

We recorded the responses of 102 V1 and 103 V2 neurons in 13 anesthetized macaque monkeys to a sequence of texture stimuli, presented in suitably vignetted 4 deg patches centered on each neuron's receptive field. The sequence, which was identical for all cells, included 20 repetitions, each of 15 samples of naturalistic and 15 samples of noise stimuli from 15 different texture families (9000 stimuli in total). The textures were each presented for 100 ms and were separated by 100 ms of a blank gray screen, so the entire sequence lasted 30 min.

2.4.1 Sensitivity to naturalistic image structure

V1 neurons responded similarly to both stimulus types, while V2 neurons often responded more vigorously to naturalistic textures than to spectrally matched noise. This distinction between V2 and V1 was evident when examining individual responses as a function of time from stimulus onset (averaged over all samples of all

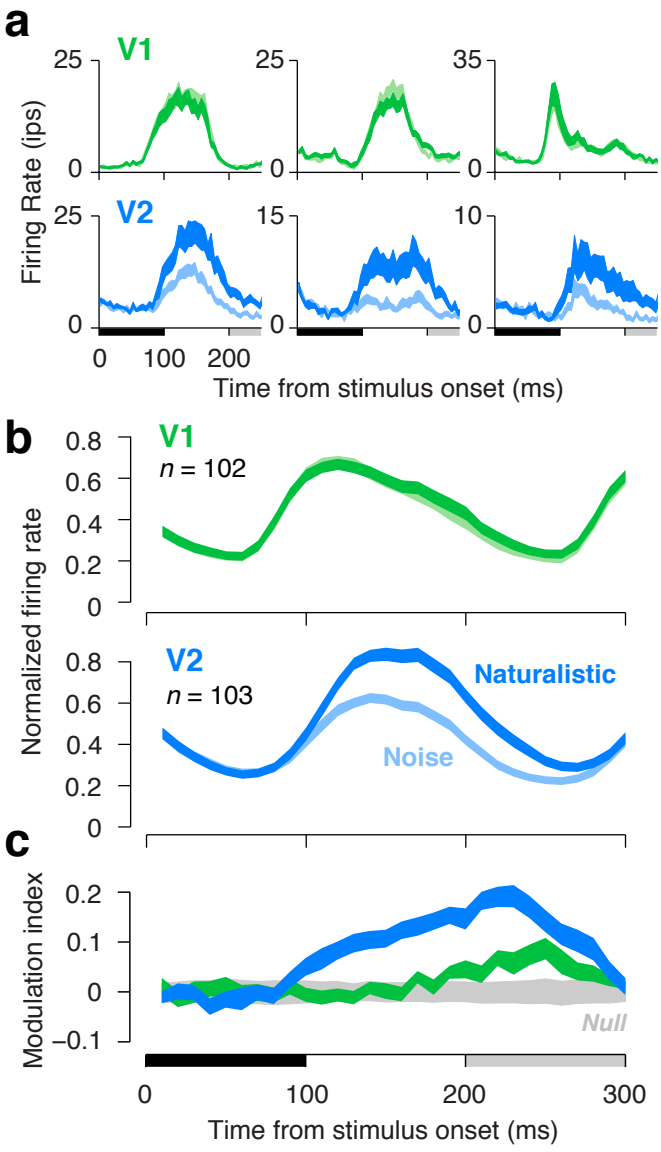

Figure 2.4: Neuronal responses to naturalistic textures differentiate V2 from V1 in macaques. (a)Time course of firing rate for three single units in V1 (green) and V2 (blue) to images of naturalistic texture (dark) and spectrally matched noise (light). Thickness of lines indicates s.e.m. across texture families. Black horizontal bar indicates the presentation of the stimulus; gray bar indicates the presentation of the subsequent stimulus. (b) Time course of firing rate averaged across neurons in V1 and V2. Each neuron's firing rate was normalized by its maximum before averaging. Thickness of lines indicates s.e.m. across neurons. (c) Modulation index, computed as the difference between the response to naturalistic and the response to noise, divided by the sum. Modulation was computed separately for each neuron and texture family, then averaged across all neurons and families. Thickness of blue and green lines indicates s.e.m. across neurons. Thickness of gray shaded region indicates the 2.5th and 97.5th percentiles of the null distribution of modulation expected at each time point due to chance.

texture families) (Fig. 2.4a), and when the responses were averaged over the cell populations (Fig. 2.4b). We use the term "modulation" to capture the differential responses to textures and noise, and index its magnitude by taking the difference of responses divided by the sum (Fig. 2.4c). The average modulation index of neurons in V1 was near zero for most of the response time course, except for a modest late positive modulation (Fig. 2.4c). Neurons in V2 showed a substantial

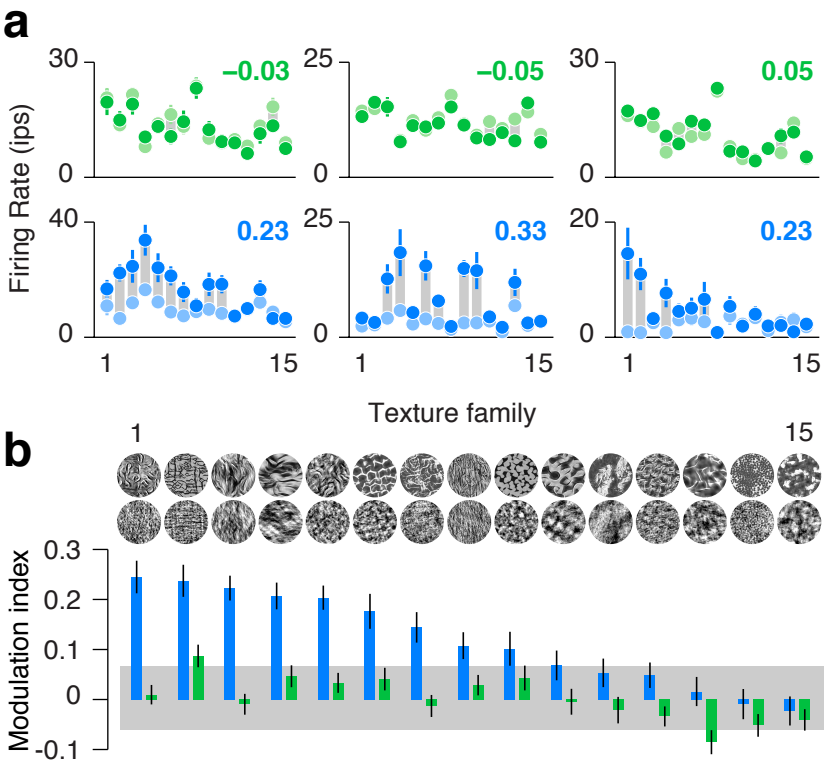

Figure 2.5: Diversity of modulation across families in V2. (a) Firing rates for three single units in V1 (green) and V2 (blue) to naturalistic (dark dots) and noise (light dots), separately for the 15 texture families. Families are sorted according to the ranking in panel b. Gray bars connecting points are only for visualization of the differential response. Modulation indices (averaged across texture families) are reported in the upper right of each panel. Error bars indicate s.e.m. across the 15 samples of each texture family. (b) Diversity in modulation across texture families, averaged across all neurons. Error bars indicate s.e.m. across neurons. Gray bar indicates 2.5th and 97.5th percentiles of the null distribution of modulation expected due to chance.

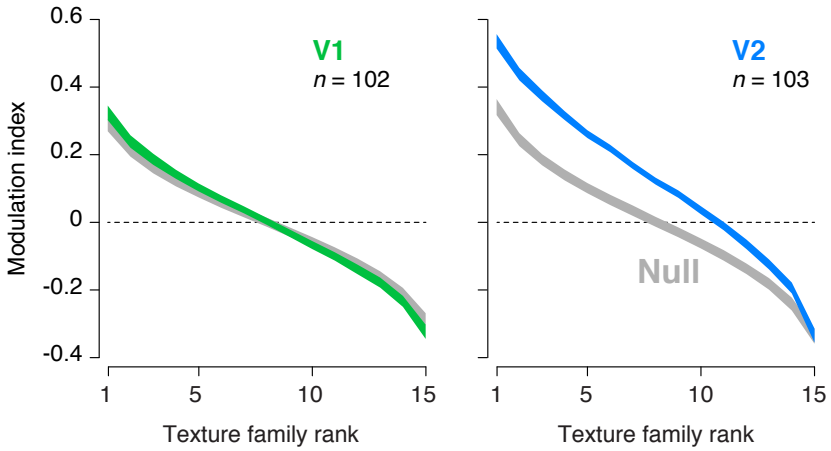

Figure 2.6: Average texture family ranking by modulation. For each cell, the 15 texture families were ranked and sorted according to modulation index. The sorted modulation indices were then averaged across all cells in V1 (green, panel a) and V2 (blue, panel b). A null distribution was obtained for both areas by permuting the naturalistic and noise labels and iterating 1000 times. Gray areas in both panels indicate the 2.5th and 97.5th percentiles of the null distribution. On average, any differences between naturalistic and noise stimuli exhibited by V1 cells, either positive or negative, were not distinguishable from those expected by chance, but this was not the case for V2 cells.

modulation that was evident soon after response onset and maintained throughout the duration of the response (Fig. 2.4c). The late modulation in V1 might reflect feedback from V2 or other higher areas [10].

V2 responses were significantly modulated by naturalistic structure on average, but the modulation was typically more pronounced for some texture families than for others. We examined responses as a function of texture family, averaged over all samples. There was a consistent trend across the V2 population for some texture families to evoke stronger modulation than others, although the most effective families varied from cell to cell (Fig. 2.5). By contrast, all families yielded negligible modulation of V1 responses (Fig. 2.5). In V2, the modulation strength across texture families was not significantly correlated with the response magnitude (r=0.42, P=0.12, correlation computed after averaging across cells). An analysis

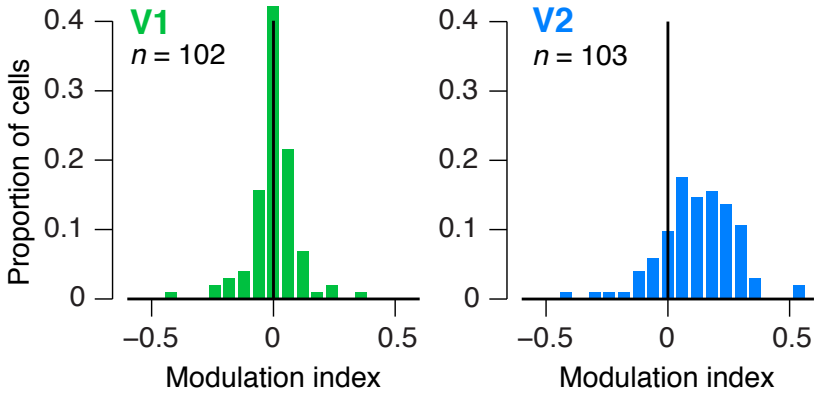

Figure 2.7: Distributions of modulation indices across single neurons in V1 and V2. For each neuron, the modulation index for each texture family was computed on firing rates averaged within an 100 ms window following response onset, and modulation was then averaged across families.

of the distribution and ranking of modulation across individual neurons ruled out the possibility that modulation in V1 was present but concealed by the process of taking means (Fig. 2.6).

Some neurons were more sensitive overall to naturalistic structure than others. We computed a modulation index for each neuron, averaged over the response duration and over all samples of all texture families (Fig. 2.7). Significant positive modulation was observed in 15% of V1 neurons, and 63% of V2 neurons ($P<0.05$, randomization test for each neuron). The difference in modulation between V1 and V2 was significant ($P<0.0001$, t-test on signed modulation; $P<0.0001$,t-test on modulation magnitude, ignoring sign). Results were similar when examining firing rates instead of the modulation index (Fig. 2.8).

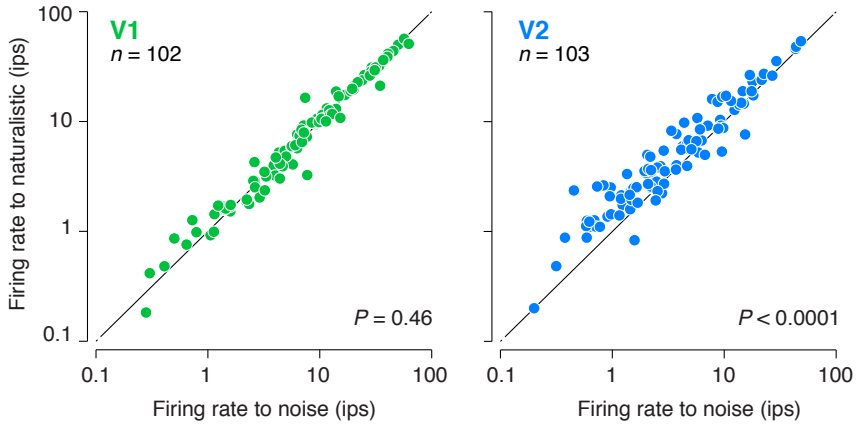

Figure 2.8: Difference between V1 and V2 in terms of firing rate. Responses to naturalistic and spectrally matched noise images in V1 (green, panel a) and V2 (blue, panel b). Diagonal dashed line is the line of equality. The difference between naturalistic and noise was statistically significant in V2 but not in V1 (paired t-tests), and the difference of differences between the two areas was statistically significant (P $<$ 0.0001, unpaired t-test). Neurons with comparable responsiveness (5-10 ips) show a differential response to naturalistic images in V2 but not in V1.

2.4.2 Relationship to receptive field properties

The receptive fields of V2 neurons are larger than those of V1, but this distinction did not explain the observed differences in sensitivity to naturalistic structure (Fig. 2.9). The stimuli presented to V1 and V2 cells were of the same size (diameter), roughly twice that of a typical V2 receptive field, and 4 times that of a typical V1 receptive field. There was no evidence for a correlation between receptive field size and modulation in either visual area (V1:r=0.13, V2:r=–0.13, P$>$ 0.05, Fig. 2.9a). When we restricted our analysis to subsets of neurons matched for average receptive field size, the difference in modulation index between areas was reduced by only 9% and remained significant (P$<$0.0001, randomization test).

We also made measurements on a subset of cells in which the stimuli were confined to each neuron's classical receptive field (CRF). In V1, the modulation was

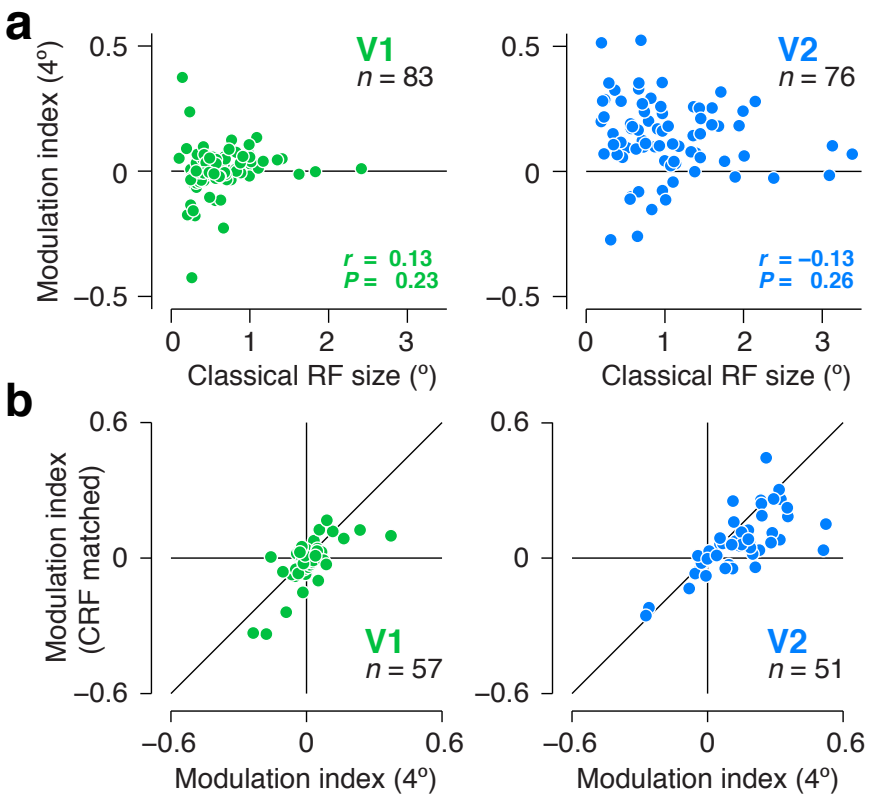

Figure 2.9: Receptive field size does not explain differential responses to naturalistic texture stimuli in V2. (a) Modulation index (ordinate) versus classical receptive field size (abscissa). Each data point represents a neuron. There was no evidence for a relationship between modulation index and classical receptive field size in either V1 or V2. (b) Comparison of modulation indices measured using stimuli presented in an aperture matched in size to the classical receptive field (ordinate) versus indices measured using stimuli presented within a 4° aperture (abscissa). Each data point represents a neuron. Diagonal dashed line is the line of equality. Modulation in V1 was near 0 for both stimulus sizes. Modulation in V2 was positive for both stimulus sizes, but there was significantly less modulation in V2 for the smaller size.

near 0 for both CRF-matched and large stimuli, though there was a small but significant reduction in modulation for the smaller stimuli (P<0.05, t-test, Fig. 2.9b). In V2, there was a robust but incomplete reduction in modulation for the smaller stimuli (P<0.0001, t-test, Fig. 2.9b), suggesting that the modulation in V2 depended partly, but not entirely, on interactions between receptive field center and surround. We found no evidence for a relationship in V2 between the modulation and commonly characterized properties of early visual neurons, including surround suppression, orientation tuning bandwidth, preferred spatial frequency, spatial frequency tuning bandwidth, or parameters of the contrast sensitivity function (c50 and exponent) (all P > 0.05, correlation). We therefore believe that our measurements reveal a hitherto unrecognized dimension of visual processing in macaque V2.

We wondered whether there was any relationship between the strength of mod-

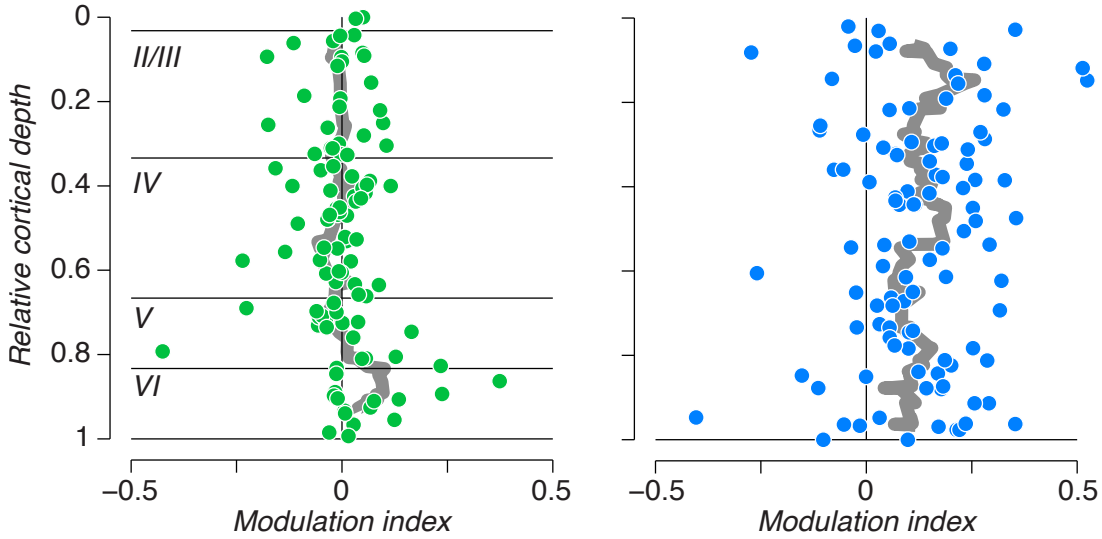

Figure 2.10: Lack of relationship between modulation and cortical depth. Each neuron was assigned a relative cortical depth based on the estimated top of layer I and bottom of layer VI in V1 and V2. Gray shaded region represents a running average over 10 neurons. V1 borders come from [195]

ulation and the architecture of visual cortex. For each of our recorded neurons we measured the depth of the electrode relative to an estimate of the top of layer I and the bottom of Layer VI. In V1, There was very little modulation at all depths of cortex, with a slight trend toward neurons with outlying positive modulation in the deep layers (Fig. 2.10). Similarly, in V2, modulation was of roughly equal strength throughout the cortical depth. We found no evidence that sensitivity to naturalistic features may be strengthened through transformations taking place between the input and output layers of V2 (Fig. 2.10; but see [256]).

2.5 Differentiating V2 from V1 in human

2.5.1 fMRI modulation across naturalistic families

Given the reliable effect of higher-order image statistics on the responses of V2 neurons, we wondered if similar effects could be observed in humans using functional magnetic resonance imaging (fMRI), which can capture large-scale differential responses across visual areas [246]. We presented alternating blocks of naturalistic and noise stimuli — one texture family at a time — in the near-peripheral visual field while measuring blood-oxygenation level dependent (BOLD) fMRI responses

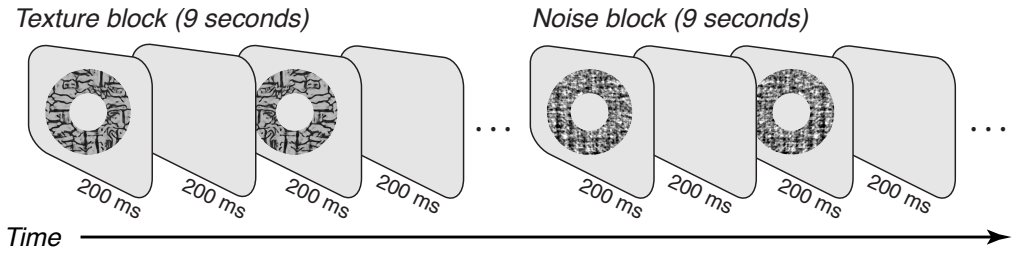

Figure 2.11: fMRI experimental design. Subjects viewed a peripheral annulus that alternated every 9 seconds between rapid presentation of spectrally matched noise and naturalistic stimuli.

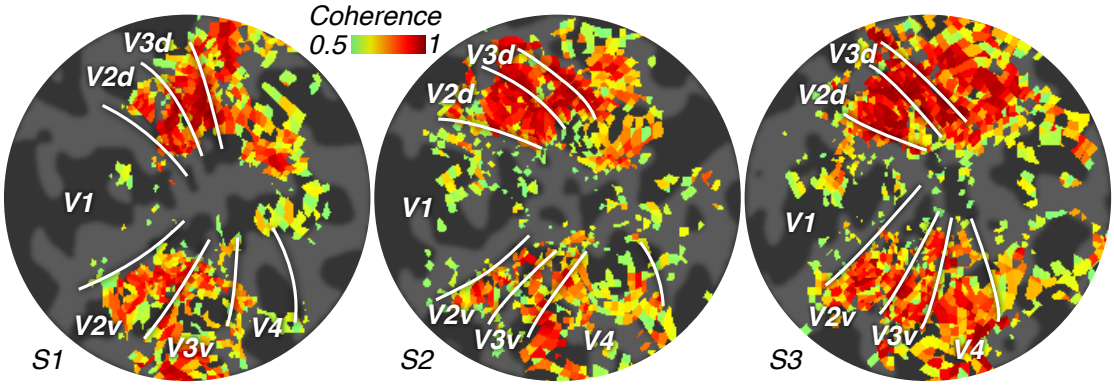

Figure 2.12: fMRI responses to naturalistic textures differentiate V2 from V1 in humans. Responses to alternating blocks of naturalistic texture images and spectrally matched noise shown on a flattened representation of the occipital pole. Color indicates coherence, which captures the extent to which the fMRI responses to naturalistic and noise stimuli differ, computed voxel by voxel after averaging responses to all texture families. White lines indicate boundaries between visual areas identified in an independent retinotopic mapping experiment.

in visual cortex (Fig. 2.11).

Subjects performed a demanding task at the center of gaze to divert their attention from the peripheral stimulus. Responses were visualized on a flattened representation of the occipital lobe, and boundaries between V1 and V2 were derived from independent retinotopic mapping [119, 246].

In all three subjects, there were strong differential fMRI responses to naturalistic texture throughout V2, and weaker ones in V1 (Fig. 2.12). We captured differential fMRI responses evoked by naturalistic texture and noise stimuli with a modulation index analogous to that used for single-unit physiology. Differences in modulation between V2 and

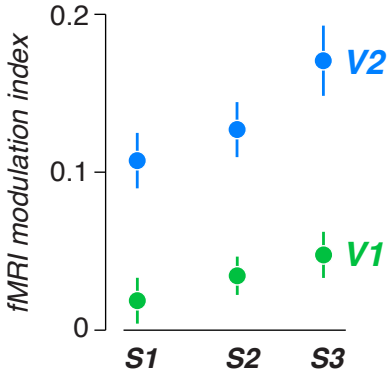

Figure 2.13: fMRI modulation index across subjects. A measure of fMRI modulation (see Methods) averaged across voxels and texture families in V1 and V2 for three subjects. Error bars indicate s.e.m. across texture families.

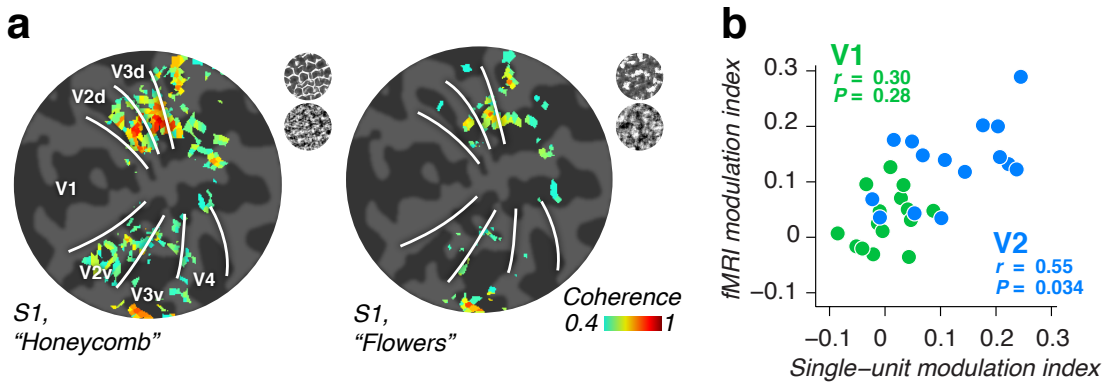

Figure 2.14: fMRI responses across different texture families. (a) Responses from a subject to two individual texture families, only one of which evoked robust differential responses in V2. Same format as Fig. 2.12. (b) Correlation between fMRI and single-unit modulation for V1 (green) and V2 (blue). Each data point represents a different texture family.

V1 were significant in each subject (Fig. 2.13; P<0.0001, paired t-test comparing responses in V1 and V2 across the 15 texture families). The much weaker modulation in V1 was nevertheless significantly greater than 0 in two of three subjects (Fig. 2.13; P<0.05). Modulation was also evident in V3, and to some extent in V4, though weaker in higher object-selective areas like the lateral occipital complex (LOC). The modulation in V3 and V4 might be inherited from V2. These results complement the single-cell findings by showing that the same response differences were evident over all of V2, and were sufficiently robust to manifest at the coarse spatial scale of fMRI.

As for the single-cell responses, some texture families elicited more robust fMRI modulation in V2 than others (see examples in Fig. 2.14a). We compared, across texture families, the fMRI and single-unit modulation indices (averaged across neurons). fMRI and electrophysiological measures were significantly correlated in V2 (Fig. 2.14b; r=0.55, P<0.05), but this was not evident in V1 (r=0.30, P=0.28). We also correlated the modulation indices from each individual neuron with the

fMRI response modulation, and found that correlations were significantly higher in V2 than in V1 ($P<0.005$, t-test on Z-transformed correlations). The presence and diversity of the differential responses to naturalistic textures in V2 are thus similar when measured in macaque neurons and human fMRI.

2.5.2 From naturalistic to natural images

We wondered how the neural representation of naturalistic textures might differ from intact natural images containing the full complexity of scenes and forms present in natural vision. Although the parameters of the texture model are largely sufficient to capture the appearance of natural images of visual texture (Fig. 2.3, [172, 16]), this is not the case for most images containing recognizable objects (Fig. 2.15, [184]). In addition to our previous experiment alternating blocks of naturalistic and noise images, we ran two additional experiments, alternating blocks of natural and noise images, as well as natural and naturalistic texture stimuli. We found a very different pattern of activation across the three alternations. As noted above, the alternation of naturalistic and noise stimuli drove the largest modulation in mid-level areas such as V2, but yielded little modulation in higher areas. In contrast, alternation between natural images and spectrally matched noise drove similar levels of modulation in V2 and V3, but much larger modulation in higher areas such as LOC (Fig. 2.16, [133]). Alternation between natural images and naturalistic textures preserved this modulation in higher areas, but nearly abolished modulation in V2 (and V3, 2.16). None of the alternations drove significant modulation in V1, consistent with the selectivity of V1 being largely driven by spectral features of images. Similarly, we observe little differential response in V2 between natural images and naturalistic textures, indicating that the statistics

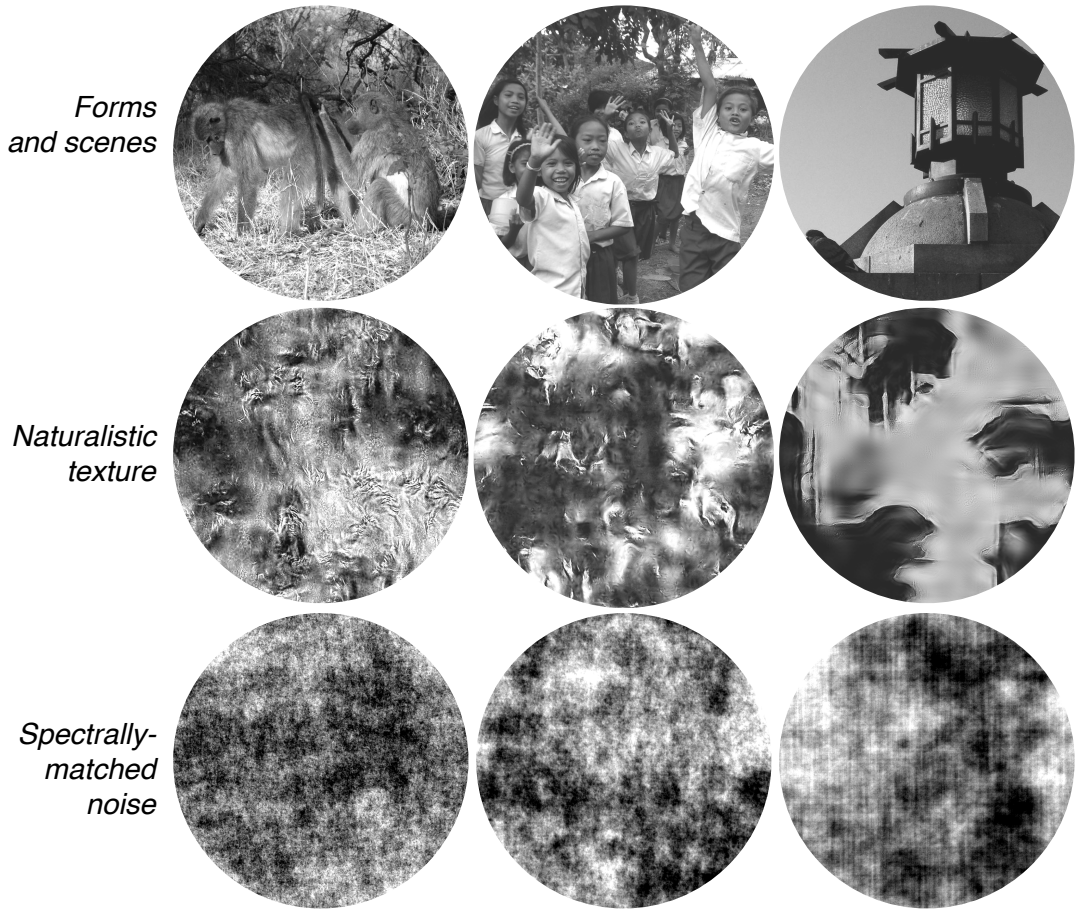

Figure 2.15: From textures to forms. Rows: Three levels of naturalness, from original photographs to naturalistic texture to spectrally matched noise. Columns: 3 example original images and their scrambled counterparts from the 30 used as stimuli.

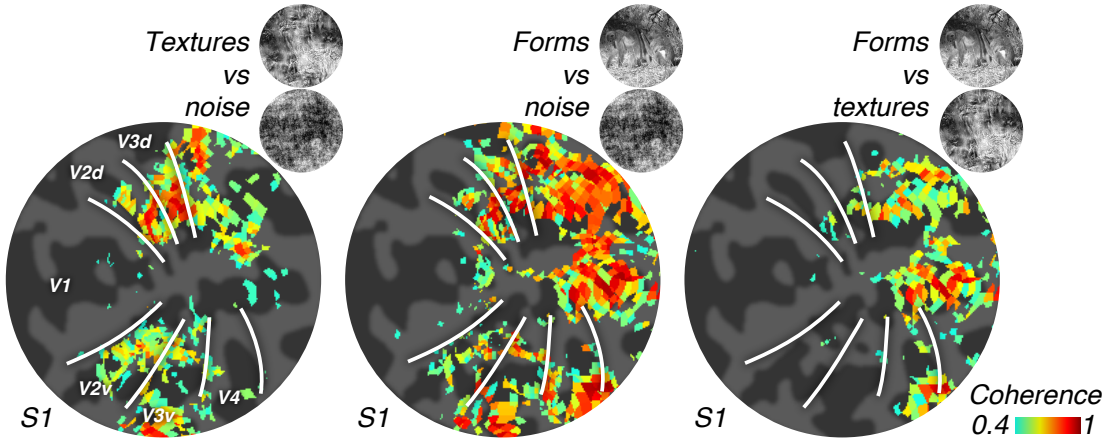

Figure 2.16: fMRI responses across three levels of naturalness. Responses from a single subject to alternations between three types of stimuli. Same format as in Fig. 2.12.

used to synthesize our texture stimuli may be sufficient to capture the selectivity of neurons in V2.

2.6 Linking neuronal and perceptual sensitivity

2.6.1 Psychophysics from the lab

If this distinctive feature of V2 responses has a perceptual correlate, then texture families that evoke larger differential responses should be those for which the naturalistic textures are more perceptually distinct from spectrally matched noise. To test this hypothesis, we built textures with varying degrees of "naturalness" (Fig. 2.17a), by titrating the inclusion of higher-order correlations in the synthesis process. We measured perceptual sensitivity to naturalness for each texture family using a three-alternative forced choice discrimination task (Fig. 2.17, 2.18) suitable for studying stochastic stimuli like textures [95].

Across 15 texture families, perceptual sensitivity was significantly correlated with electrophysiological response modulations, averaged across neurons in V2

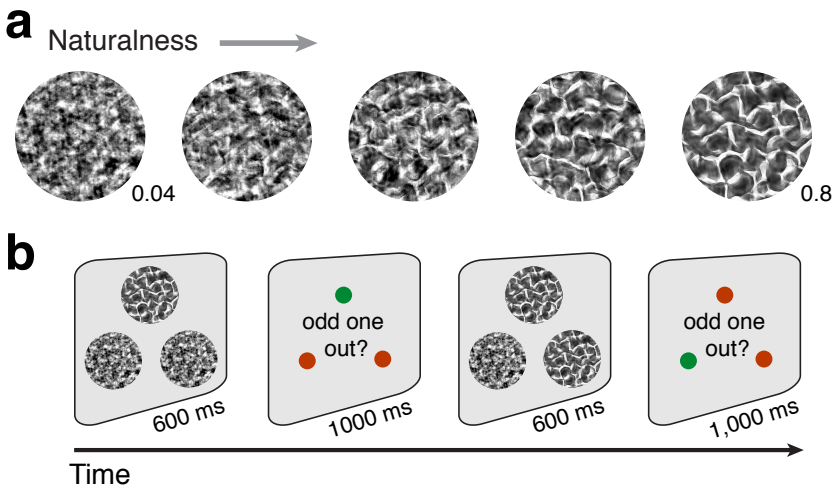

Figure 2.17: Psychophysical task. (a) Stimuli were generated along an axis of "naturalness" by gradually introducing higher-order correlations (Fig. 2.1c). (b) Observers performed a 3AFC "oddity" task in which they viewed three images, two naturalistic and one noise (or vice versa), and indicated which looked different from the other two. All three images were synthesized independently (e.g., starting with statistically independent samples of Gaussian white noise).

(Fig. 2.19a; r=0.62, P<0.05), but not in V1 (r=0.21, P=0.45), and the correlation was significantly larger for V2 than V1 (P<0.0001, t-test on Z-transformed correlations). We also found that perceptual sensitivity was significantly correlated with

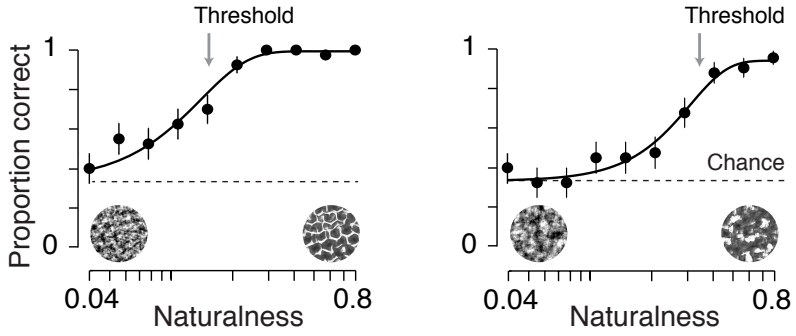

Figure 2.18: Example psychometric functions: performance as a function of naturalness. Solid curves, best-fit cumulative Weibull function. Chance performance is 1/3. The two panels show two different texture families (same as in Fig. 2.14a) with different thresholds (defined as the level of naturalness required to obtain 75% correct).

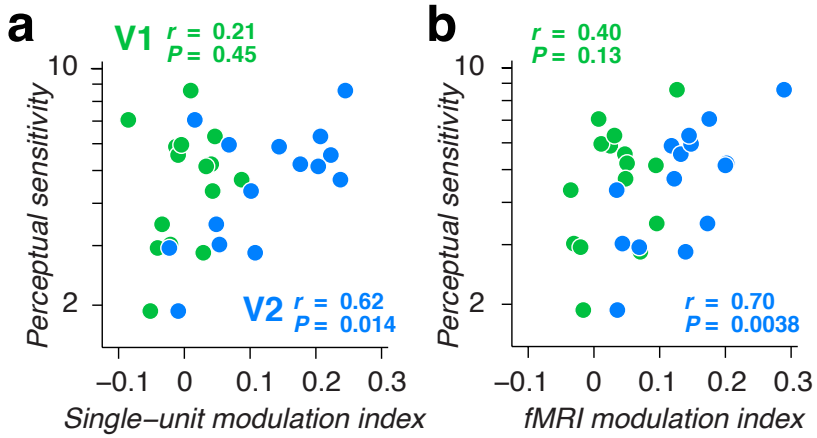

Figure 2.19: Neuronal responses to naturalistic textures in V2 predict perceptual sensitivity. (a) Correlation between psychophysical sensitivity (1/threshold) and single-unit modulation in V1 (green) and V2 (blue). Each data point represents a texture family. (b) Correlation between psychophysical sensitivity and fMRI modulation. Same format as panel a.

the fMRI modulation in V2 (Fig. 2.19b; r=0.70; P<0.005) but not in V1 (r=0.40; P=0.13), and that this correlation was again significantly larger for V2 than V1 (P<0.01, paired t-test on Z-transformed correlations). These relationships suggest a functional role for V2 in the perception of these naturalistic stimuli.

2.6.2 Psychophysics from the crowd

The texture families we used varied in the form and extent of their statistical dependencies. We wondered which of the many possible dependencies were most important for perception and — by extension — for evoking responses in V2. Identifying the relevant subset requires a large number of stimuli, but making biological measurements – neuronal or fMRI – for such an ensemble would be unfeasible. We therefore measured perceptual sensitivity for nearly 500 texture families using Amazon.com's Mechanical Turk service to "crowdsource" our measurements [164] and expand their range 30-fold. This approach yielded a total of 300 hours of

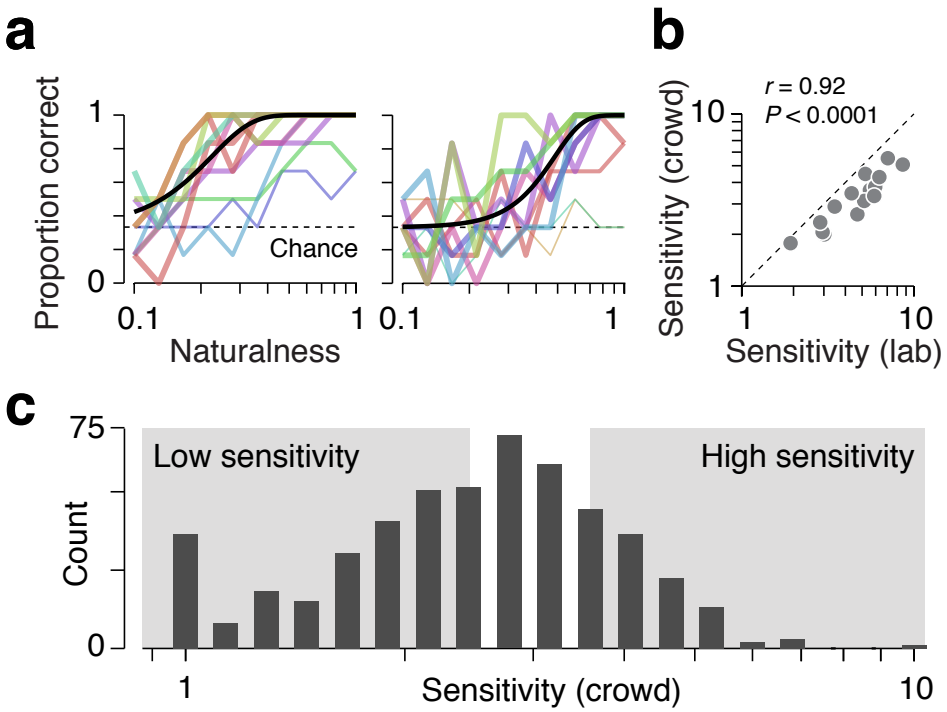

Figure 2.20: Crowdsourced psychophysical estimates of sensitivity for hundreds of texture families. (a) Example psychometric functions for two texture families (same as Fig. 4c and 5c), each based on observers recruited from Amazon.com's Mechanical Turk performing a 3AFC task in a web browser. Each colored line corresponds to one observer. The black line indicates the best-fitting psychometric function, estimated using a mixture model that re-weighted observers based on their reliability; thickness of the colored lines indicates the weight assigned to each observer. Chance performance is 1/3. (b) Perceptual sensitivity (1/threshold) was significantly correlated when measured in the laboratory (abscissa) and in the crowd (ordinate). Dashed line is the line of equality. (c) The distribution of perceptual sensitivities across 494 texture families was used to pick 20 families spanning the range of sensitivities, emphasizing the extremes (light gray regions).

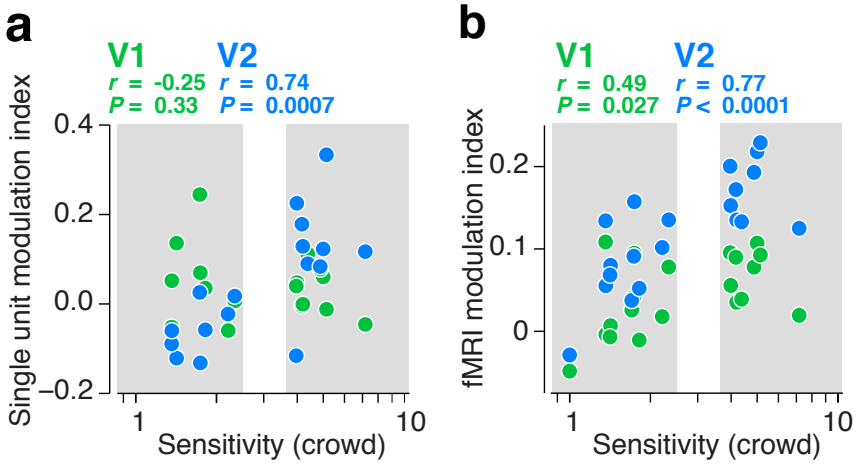

Figure 2.21: Crowdsourced psychophysical estimates predict V2 responses. (a) Correlations between single-unit modulation and sensitivity (measured in the crowd) for the chosen families, in V1 (green) and V2 (blue). Only 17 of the 20 families were included due to experimental time constraints. (b) Correlations between fMRI modulation and sensitivity. All 20 families were included. Same format as panel d.

behavioral data from thousands of human observers (Fig. 2.20, see Methods). We developed analysis procedures to combine data from this large number of observers, and to evaluate the reliability of the results.

We related our crowdsourced measurements to our previous results in two ways. First, we confirmed for the original 15 texture families (Fig. 2.19) that perceptual sensitivity measured in the crowd was reliably correlated with, albeit lower than, sensitivity measured in the laboratory (r=0.92,P<0.0001,Fig. 2.20b). Second, we used the 494 new texture families to link the crowdsourced sensitivity estimates back to physiological responses. We selected 20 texture families spanning a range of crowd-estimated sensitivities, emphasizing the extremes (Fig. 2.20c). We used images from these families as stimuli in additional single-unit and fMRI experiments. Both the single-unit modulation in V2 (r=0.74,P <0.001, 16 cells) and fMRI modulation in V2 (r=0.77, P<0.0001, 2 subjects) were significantly corre-

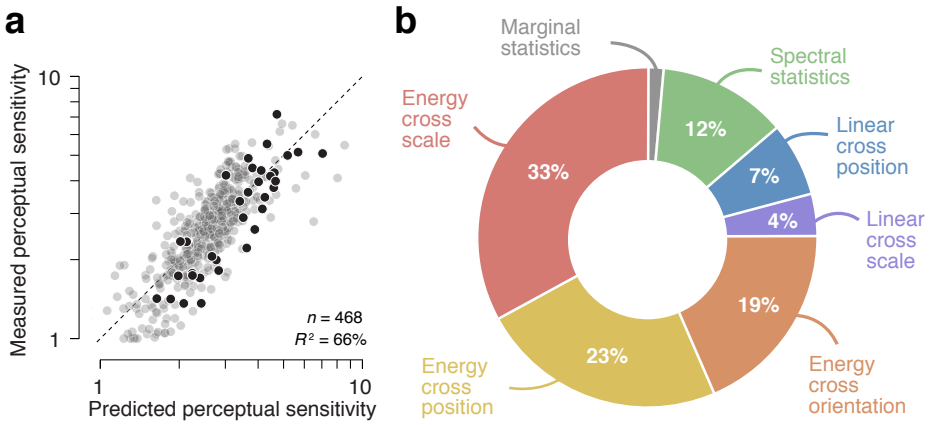

Figure 2.22: Using higher-order correlations to predict perceptual sensitivity. (a) We used multiple linear regression to predict perceptual sensitivity to naturalistic textures based on higher-order correlations and other image statistics used in texture synthesis. Each data point corresponds to a texture family; black dots indicate all texture families used in physiological experiments. Black dashed line is the line of equality. (b) Wedges indicate the fractional R^2 assigned to each group of texture synthesis parameters from the regression analysis. See [172] and [16] for example images demonstrating the role of some of these parameters in texture synthesis.

lated with crowd-estimated sensitivity (Fig. 2.21), confirming with novel stimuli the relationship found in our earlier experiments (Fig. 2.19). In V1, single-unit modulation showed no evidence for a correlation with sensitivity (r=–0.25, P=0.33, 11 cells). fMRI modulation in V1 to these new stimuli revealed a significant correlation (r=0.49, P<0.05). This was weaker than the correlation found for V2, and similar to our results using the original 15 textures (Fig. 2.19b).

The crowdsourced psychophysical data for the complete ensemble of texture families allowed us to identify which statistical dependencies of the images explained diversity in perceptual sensitivity to naturalistic structure. Recall that our textures were synthesized to match correlations among V1-like responses — both linear filter responses, and energies — at different orientations, positions and scales. Through a combination of principal components analysis and multiple lin-

ear regression (see Methods), we used these correlations, along with spectral and marginal statistics, to predict more than half of the variance in perceptual sensitivity (Fig. 2.22a, R^2=66%). To ensure that results were not a result of overfitting, we confirmed that accuracy was still high (R^2=60%) with 10-fold cross-validation. To identify the relative importance of different synthesis parameters, we decomposed the total R^2 using the "averaging over orderings" technique (see Methods) [77]. The largest share of variance was accounted for by the cross-scale correlations among the energy filter responses; second and third most important were the cross-position and cross-orientation energy-filter correlations (Fig. 2.22b). Correlations among linear filter responses were less important. Spectral properties had a small amount of predictive power, but this likely reflected how spectra control visibility, e.g., insensitivity to high spatial frequencies. The contribution of marginal statistics (skewness and kurtosis) was negligible, indicating that perceptual sensitivity is driven by the higher-order correlations rather than basic image properties. Together, these results link perceptual sensitivity — and, we infer, neuronal sensitivity in V2 — to the particular kinds of higher-order statistical dependencies found in our visual textures.

2.7 Towards a functional model

How might individual V2 neurons represent these statistical dependencies? We do not intend to suggest that V2 neurons directly encode the correlations in V1-like afferents contained in our texture synthesis algorithm [172]. Instead, a variety of nonlinear computations, similar in function but differing in detail, can effectively capture the same information and could enable the sensitivity to naturalistic stimuli that we found in V2. For example, selectivity for different kinds of correlations

could be achieved by combining squared and spatially pooled linear combinations of pairs of V1 inputs, analogously to the way 'motion energy' computations can express the correlations of the Reichardt model [4]. By indirectly computing correlations, such a mechanism can produce increased responses to naturalistic stimuli [58]. However, we wondered whether a similar framework could account for additional features of our physiological results, such as the pattern of modulation strength across families and the V1-like tuning of V2 neurons for simpler stimuli such as sinusoidal gratings. We specifically aimed to simulate our V2 results while limiting the model construction to assumptions derived from known canonical computations commonly used to model V1 neurons.

We first constructed a model V1 stage to demonstrate that standard models of early visual cortex are insufficient to account for sensitivity to naturalistic image structure. We built a bank of simulated V1 complex cells whose selectivity tiled the Fourier domain with 6 spatial frequency bands and 8 preferred orientations, and tiled space with 11 center positions in both dimensions (Fig. 2.23a). The output of each filter was normalized by the sum of the activity of all filters over the entire image patch (Fig. 2.23b). The strength of this normalization was controlled through a scale parameter that ranged from 0 to 1. To simulate our physiological V1 population we sampled the responses of 100 units selective for different orientations, spatial frequencies, spatial positions, and with different normalization constants. We generated responses to each of the 450 images (15 samples from 15 families for both noise and naturalistic) used in our initial physiological recordings and computed the average modulation index for each unit (Fig. 2.23c). We found the simulated distribution was similar to that observed for V1 neurons (Fig. 2.7).

We next constructed a second stage by taking a linear combination of differently

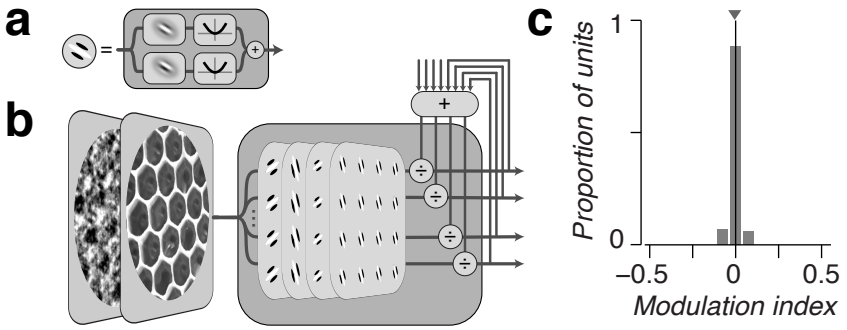

Figure 2.23: Model V1 neurons are not sensitive to higher-order structure. (a) Each V1 unit was modeled as the sum of the squared output of two linear filters in quadrature phase. (b) Signal flow diagram for V1 model. Each image was represented as the output of a population of V1 complex cells whose receptive fields tiled space, orientation, and spatial frequency. The output of each V1 cell was normalized by the sum of the total activity in the population. (c) Distribution of the modulation index over a population of 100 simulated V1 complex cells. Each unit was chosen randomly to have a specific preferred orientation, spatial frequency, position, and normalization constant. Modulation index was computed identically to that used for physiological characterization. Arrow indicates median modulation index across the population. Compare with 2.7.

tuned V1 neurons. This represents an approach to V2 that has been investigated previously both explicitly [250, 238, 239] and implicitly by positing V2 neurons combine multiple orientations within their receptive fields [103, 106, 11]. However, we lacked enough data to fit such a model to the responses of single neurons and thus had no way to constrain the linear filter applied to V1 inputs. Instead, we made the assumption that the filtering operation in V2 likely resembles the linear filters found in V1. Adelson and Bergen [6] suggested that the two-dimensional, Gabor-like filters found in V1 can be usefully considered as oriented derivatives applied to the two dimensions of space. Just as V1 neurons are selective for changes in luminance across space (as represented by their LGN afferents), V2 neurons may be selective for changes in the activity of V1 populations across space, as well as across orientation and spatial frequency.

We instantiated this idea by taking a derivative across the four dimensions

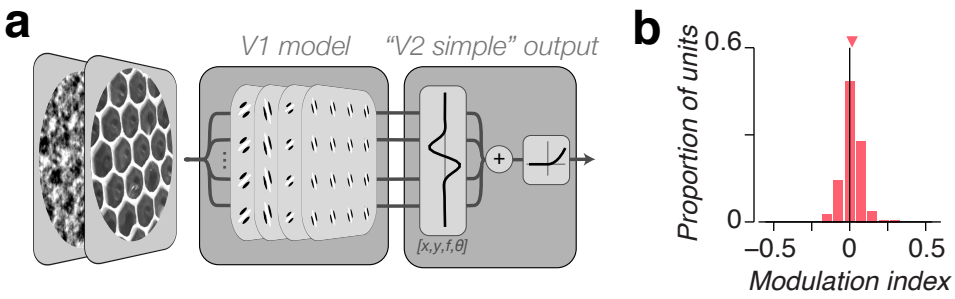

Figure 2.24: "V2 simple" cells take linear combinations of V1 afferents. (a) Model architecture for "V2 simple" cells. (b) Distribution of the modulation index over a population of 100 simulated "V2 simple" cells. Each unit had random center and orientation in the four dimensions of space, orientation, and spatial frequency. Arrow indicates median modulation index across the population. Compare with 2.7.

of simulated V1 activity. Applying a weight to the derivative of each dimension results in an oriented gradient. We simulated the response of individual "V2" units as the rectified and squared output of this derivative operation (Fig. 2.24a). We sampled the responses of 100 units to our stimuli, each with random weights applied to the derivatives across the V1 population. We found this population had a modulation index that was modestly higher than our simulated V1 population, and also exhibited more variability (Fig. 2.24b). The lack of strong modulation is consistent with previous results modeling V2 responses as linear combinations of V1 afferents [238, 239]. This previous approach to V2 did not constrain the weights on different V1 filters as we do here, allowing for much more flexibility in the V2-stage filtering operation. Even so, model responses fit directly to V2 neuronal responses failed to predict increased sensitivity to naturalistic image structure [239].

This model architecture can be considered as describing a "V2 simple" cell. Similar to V1 simple cells, the output of a "V2 simple" cell is determined by a linear combination of its afferents followed by a nonlinearity. Such an architecture can achieve weak sensitivity to naturalistic correlations depending on the form

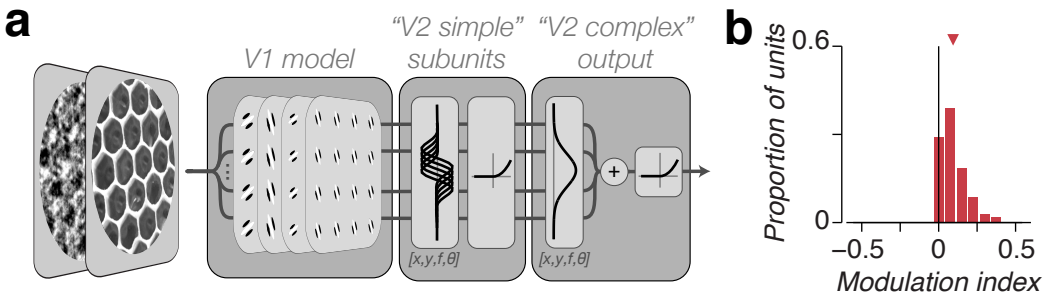

Figure 2.25: "V2 complex" cells are sensitive to naturalistic structure. (a) Model architecture for "V2 complex" cells. (b) Distribution of the modulation index over a population of 100 simulated "V2 complex" cells. Each unit had random center and orientation and pooling width in the four dimensions of space, orientation, and spatial frequency. Arrow indicates median modulation index across the population. Compare with 2.7.

of the nonlinearity. Simple rectification will not produce modulation on average [58], but the half-squaring operation we use here does induce weak modulation (because of the multiplicative term in the output of the squared difference of two V1 afferents). However, the strength of this modulation in our simulation does not approach that observed in the physiology (and increasing the exponent of the nonlinearity further did could not make up this discrepancy). We therefore constructed "V2 complex" cells that pool the output of multiple "V2 simple" cells over space, orientation, and frequency. Such a mechanism is analogous to those proposed to compute motion energy in V1 [4] or relative disparity in V2 [220].

Specifically, we convolved the output of our V1 front end with a single derivative subunit across space, orientation, and frequency (Fig. 2.25a). The output was rectified and squared and then pooled in a second stage. This second stage thus receives the output of several "V2 simple" cells that apply the same filtering operation to different combinations of V1 afferents. We varied the width of this second stage pooling (across space, orientation, and frequency) to determine how "complex" the resulting "V2 complex" cell is. If the pooling is small, the unit

will act mostly as a "V2 simple" cell; if the pooling is large, the unit will more effectively compute stimulus correlations. This convolutional framework reflects recent models of V1 computation that efficiently and elegantly capture features of V1 computation (such as the fundamentally noncategorical distinction between simple versus complex V1 cells) [187, 130, 240]. We again simulated a population by sampling units with random parameters determining the derivative orientation of the subunits and the width of a Gaussian pooling function applied to each of the four dimensions of V1 activity. We found that a population of such "V2 complex" cells contains strong modulation on average (Fig. 2.25b), qualitatively consistent with our physiological results (Fig. 2.7).

Intriguingly, we also found that this simulated population of "V2 complex" cells also accounted for the variability in modulation across texture families (Fig. 2.26). While there was no relationship between the average "V2 simple" cell modulation with V2 physiology across families (Fig. 2.26a), the average modulation of "V2 complex" cells was significantly correlated with the average V2 physiology (Fig.

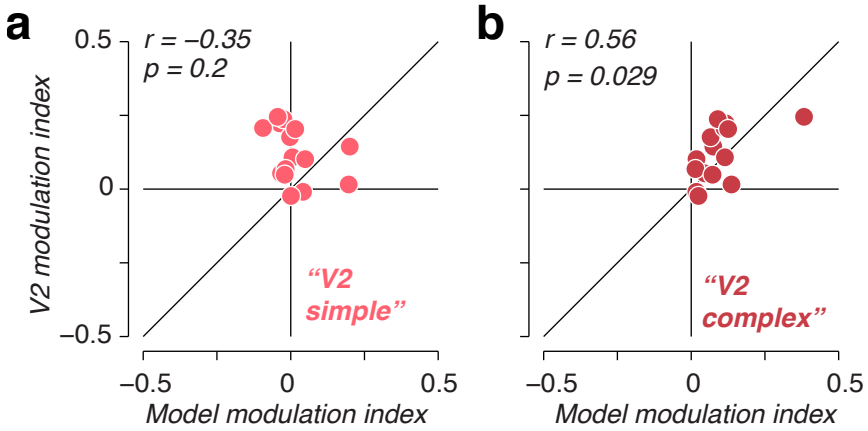

Figure 2.26: "V2 complex" cells predict physiolgical modulation across families. (a) "V2 simple" cell model and physiological modulation index for each of 15 texture families. (b) Same as in (a) but for "V2 complex" cells.

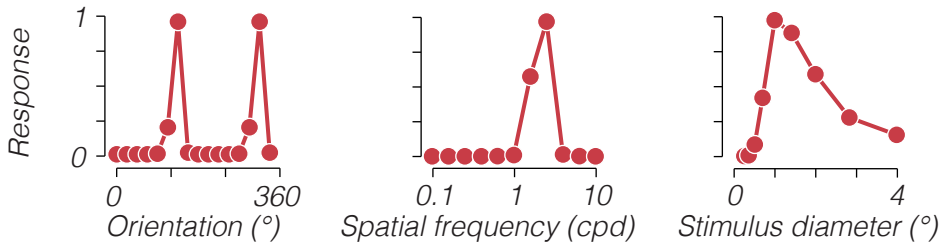

Figure 2.27: "V2 complex" cells have V1-like tuning to gratings that differ in orientation (left), spatial frequency (center) and diameter (right). This model unit had a strong modulation index of 0.19 but did not show heterogeneity in it's tuning to orientation or spatial frequency. The suppressed response to large stimuli of large diameter (a common feature observed in both V1 and V2) is inherited from the normalization in the V1 stage. Compare these tuning curves with those in Fig. 1.6.

2.26b). We also found that even model "V2 complex" cells selective for naturalistic structure often exhibit V1-like tuning to simpler stimuli such as drifting gratings (Fig. 2.27, compare with Fig. 1.6). Because the mechanism computing correlations in our model does so through pooling the squared output differences of V1 neurons that neighbor each other in space, orientation, and/or spatial frequency we do not see a large increase in tuning bandwidths or receptive field heterogeneity. These results suggest an explanation for why distinctive selectivity in V2 has been difficult to detect using the kinds of artificial stimuli often used previously [169, 89, 91, 120, 132, 106, 11, 196]. Finally, the idea of a "V2 complex" cell is conceptually satisfying because it suggests that nonlinear computations of identical form reappear at multiple stages of the cortical hierarchy [87, 33], and could be further explored in V2 by measuring responses to naturalistic or artificial stimuli containing specific higher-order correlations and predicting their responses with hierarchical models [185, 241].

2.8 Discussion

We have discovered that naturalistic texture stimuli modulate the responses of neurons in area V2, while having only a minimal effect on neurons in area V1. These modulations were similar and substantial in both anesthetized macaques and awake humans. The diversity of modulation across different texture families predicted the perceptual salience of their naturalistic structure. We capitalized on this diversity to reveal the importance of correlations across scale, and to a lesser extent across position and orientation, for driving V2 activity. The combination of human and monkey physiology, and crowdsourced psychophysics, provide mutually reinforcing evidence that V2 plays a direct functional role in representing naturalistic structures.

Previous studies have identified specialized response properties for representing visual form in subpopulations of V2 neurons [106, 11, 250], but the differences between V2 and V1 were usually small [169, 120, 132, 91, 196]. Some of these may reflect special cases of the properties identified here, such as tuning for angles reflecting sensitivity to cross-orientation correlations. The attribute that has most robustly distinguished V2 from V1 is "border ownership" [259], which may also depend on the receptive field surround in V2 [40, 69]. Border ownership signaling, however, may rely on attentional feedback [176, 48], whereas the response pattern we have discovered probably does not, as it is evident in both awake humans with diverted attention and anesthetized macaques.

Our fMRI measurements robustly differentiated human V2 from V1. However, unlike in our single-unit recordings, there was a weak, significant correlation between fMRI measurements in V1 and perceptual sensitivity (Fig. 2.19b, 2.21b).

These V1 signals may reflect the influence of modulatory feedback [10]. Such an influence was hinted by the late component of modulation in the V1 single-unit response time course (Fig. 2.4c), and could be more readily evident with fMRI [177]. Establishing a more direct relationship would require further study of the late V1 single-unit response — by recording from more neurons, and thus more reliably measuring the weak signal, or employing techniques, like cooling, capable of isolating or disabling feedback signals [55, 146].

We compared responses to naturalistic texture stimuli with responses to spectrally matched noise images, similar to the globally phase-randomized images that have been used previously in fMRI [133], psychophysics [221], and physiology [54] experiments. Presentation of intact and phase-randomized objects, for example, reveals differential fMRI responses throughout the human lateral occipital cortex [133]. But none of these studies reported differences between V1 and V2. This may be due to the use of uncontrolled images of natural objects or scenes [53, 186], which obscures the influence of the higher-order statistical dependencies upon which we have focused, and instead emphasizes responses in downstream object-selective areas. A previous study of V1 and V2 used natural photographs as stimuli [250], but this study had different goals and did not relate neuronal responses to the particular statistical dependencies considered here. The spatial homogeneity of our stimuli, coupled with a synthesis method that enforced a particular set of higher-order statistical dependencies, facilitated robust and specific links between neuronal responses in V2, image statistics, and perception. Our ability to generate multiple images from each texture family also facilitated comparisons between neurophysiology (averaging across neurons with different receptive field locations) and human fMRI [86]. Synthetic naturalistic stimuli like ours thus offer a balance be-

tween natural and artificial that may prove useful in physiological characterization in other sensory domains [137, 136].

We used a large-scale, crowdsourced psychophysical experiment to show that particular subsets of higher-order statistical dependencies predicted diversity in perceptual sensitivity — and, by extension, neuronal responses in V2. Correlations among energy filter responses (akin to V1 complex cells) were more important than among linear filter responses (akin to V1 simple cells), which is notable given that V2 neurons receive input from both simple and complex cells [195]. The particular computation implied by the responses to our stimuli may depend primarily on complex cell input. This hypothesis could be further explored by combining our stimulus protocol with measures of V1-to-V2 connectivity[72]. Our analysis of crowdsourced data also revealed the importance of dependencies across scale, followed by dependencies across position and orientation. Most studies of V2 thus far have emphasized interactions across orientation, e.g., by measuring responses to curvature or angles[89, 91, 106, 11, 238]. These visual elements are salient in man-made environments, but may play an outsized role in our intuitions about how the visual system begins the process of parsing natural scenes. Instead, we infer that V2 neurons might be particularly sensitive to dependencies across scale, which are equally fundamental to natural image structure.

The transformation of visual information as it ascends the cortical hierarchy enables the perception of scenes and objects. A common view is that early computations encode the primitive elements of which scenes are made, and that subsequent stages of processing assemble these elements into larger and more complex combinations, capturing the structural relationships that determine the visual world. This constructionist view has stumbled on the problem of V2, whose neurons have

stubbornly refused to reveal the form of their preferred elementary feature combinations [169, 89, 91, 120, 132, 106, 11, 238, 196]. We have found it useful to attack this problem with well-controlled texture stimuli that emphasize the statistical regularities of natural images, as well as with stimuli containing more conventional visual features. Our findings suggest that two fundamental constituents of visual scenes — the specific feature combinations that comprise objects, and the statistics that define textures — may both be represented in V2 [5].

Chapter 3

Selectivity and tolerance in V2

3.1 Introduction

Visual perception in primates arises from the responses of neurons in a variety of areas within the cerebral cortex. These responses are typically characterized by measuring selectivity for specific visual attributes, such as light intensity or color, and local structural properties such as spatial position, orientation, and spatial frequency. Stimulus selectivity, along with the complementary notion of "invariance" or "tolerance" to irrelevant variation, provides a de facto language for describing the functional roles and relationships of neurons in visual areas. For example, simple cells in area V1 are selective for orientation [101] and spatial frequency [143, 224, 42]. Complex cells exhibit similar selectivity, but are also more tolerant to changes in spatial position [101, 141, 4]. Component cells in area MT (V5) exhibit selectivity for orientation and speed, but (relative to their V1 inputs) are more tolerant of changes in location and spatial frequency, whereas MT pattern cells are tolerant to changes in orientation (and more generally, spatial structure) [144].

Neurons in area IT are selective for visual images of particular objects, but are

tolerant to identity-preserving transformations, such as translation, rotation, or background context [263, 242]. This tolerance increases from area V4 to IT [184], suggesting that an increase in selectivity is balanced by an increase in tolerance, preserving overall response levels and their distribution across neurons [183]. But the selectivity and tolerance of visual representations in midventral areas, in particular V2, has been more difficult to establish because of the lack of knowledge of the relevant visual attributes. V2 neurons receive much of their afferent drive from V1, have receptive fields that are roughly twice the size of those in V1, and exhibit similar selectivity for orientation and spatial frequency [64, 123]. Indeed, the responses of V2 neurons to many forms of artificial stimuli, including gratings, curves, and texture-defined patterns are only modestly different than the responses of neurons in V1 [106, 91, 11, 196].

Recent work suggests that local statistical measurements that capture the appearance of visual textures might provide a feature space for characterizing the responses of V2 neurons [60, 61, 256]. Sensitivity to multipoint correlations in arrays of binary (black and white) pixels first arises in V2 [256], and is strongest for those correlations that are most informative about binarized natural images [222], and that are most perceptually salient [235]. This sensitivity to higher-order correlations is also present for more naturalistic stimuli. Images of natural visual texture evoke correlated responses in rectified V1-like filters tuned for differing orientation, scale, and position [172]. V2 neurons are well-driven by synthetic texture stimuli containing these naturally occurring correlations, and less so by texture stimuli that lack them [61]. Moreover, the performance of human observers in detecting these correlations is predicted by the differential increase in average V2 response levels [61]. All of these results provide evidence that area V2 plays a role in rep-

resenting the higher-order statistics of visual textures, but the ways in which this representation supports visual tasks such as discrimination have yet to be explored.

Here, we provide a more direct test of the link between V2 and the representation of the higher-order statistics of natural textures. We generated stimuli that are matched to the statistics of naturally occurring homogeneous texture images. These stimuli are perceptually similar to one another, and similar to the original texture image, despite marked differences in the position and detailed arrangement of their local features [172, 16, 1]. This property can be used to generate pronounced distortions in peripheral viewing, which remain imperceptible so long as the distortions preserve texture statistics over spatial regions the size of V2 receptive fields [60]. If V2 is encoding these local statistics, and is responsible for these perceptual phenomena, then the responses of populations of V2 neurons to statistically matched stimuli should reveal a particular form of tolerance. Specifically, populations of neurons in V2 should respond similarly to stimuli that are statistically matched, despite variation in local image detail. This kind of tolerance would complement previously reported tolerances to geometric image transformations, such as translation or rotation, found at higher levels of visual cortex [263, 242, 184].

We studied this tolerance to statistical resampling by analyzing responses of a collection of V1 and V2 neurons to images of synthetic texture, generated to match the statistics of different texture "families." V2 responses across families of statistically matched stimuli were more homogeneous than V1 responses, reflecting an increased tolerance that was only partly explained by the larger size of their receptive fields. Using a neural population decoder, we found V2 was better than V1 at discriminating between-family images matched for different statistics and

worse at discriminating within-family images matched for the same statistics, a pattern of performance that broadly resembles human perceptual experience [172, 1].

3.2 Methods

Stimuli and physiological methods for this chapter were identical to those reported for chapter 2 in sections 2.2.1 and 2.2.2.

3.2.1 Analysis

ANOVA

For all quantitative analyses, we averaged spike counts within a 100 ms time window aligned to the response onset of each single unit. Response onset was determined by inspection as the first time point eliciting a response above baseline; results were nearly identical when using a quantitative criterion based on the standard deviation of the response. We first applied a Freeman-Tukey variance-stabilizing transformation, $\sqrt{x+1}+\sqrt{x}$ [62], to the spike counts for each neuron. We then performed a nested analysis of variance (ANOVA) to partition the total variance into the portions arising across families, across samples within a family, and across repetitions of the same stimulus. The ANOVA generates an F-statistic that captures the ratio of variances between each hierarchical level. For the vast majority of neurons, the F-statistic was significant for ratios of variance across repetitions and across samples (101/102 in V1, 103/103 in V2) as well as for ratios of variance across samples and across families (91/102 in V1, 97/103 in V2). We chose to perform further analysis using the ratio between partitioned variance, but

all results were qualitatively similar when using the F-statistic itself. To obtain the variance ratio we divided the percent variance across families by the percent variance across samples. To avoid outlying values when either variance was very low, we stabilized the ratio by adding 2% variance to both the numerator and denominator. We tested how reliable our estimates of the variance ratio were by splitting the 20 repetitions for each condition in half and performing the ANOVA separately on both halves of the data for each neuron. We repeated this process 10,000 times with different partitions of the original repetitions and asked how well our estimate on one half of the data could predict the other half.

Regression

Basic receptive field properties for each neuron — e.g., receptive field size, contrast response function — were determined offline by using maximum likelihood estimation to fit an appropriate parametric form to each tuning function. These fits were only obtainable for a subset of neurons (84% in V1, 73% in V2) due to incomplete characterization arising from time constraints during the experiment. We first asked how well we could predict the log variance ratio in each area using a large number of receptive field properties (preferred spatial frequency, spatial frequency bandwidth, orientation selectivity, classical receptive field size, contrast exponent, c50, maximum firing rate, surround suppression index, modulation ratio (f1/f0), and texture modulation index (see 2.7). We used the log variance ratio because the ratios were approximately normally distributed in the log domain. We used a stepwise linear model to estimate which receptive field properties added to the goodness of fit. For V1, only receptive field size and modulation ratio were included in the model. For V2, receptive field size and modulation ratio

were included along with orientation selectivity and semisaturation contrast (c50). Classical receptive field size was defined as the standard deviation of the center in a ratio of Gaussians model. The modulation ratio was computed from responses to one-second presentation of an optimal grating, and represents the ratio between the first harmonic and mean of the average response. The orientation selectivity index (OSI) was computed as the circular variance of the baseline-subtracted firing rates to each orientation, so that OSI = 0 indicated no selectivity and OSI = 1 indicated sharp tuning for orientation. Semisaturation contrast (c50) represents the contrast level that evoked half the maximum firing rate in a Naka-Rushton fit to the responses to a grating of varying contrast. To examine how each of these predictors contributed to the variance ratio we used an averaging-over-orderings [77, 61] technique to estimate variance explained by each receptive field property. This allowed us to asses the relative importance of each predictor in each area. We computed error bars for the contribution of each receptive field property and the overall explained variance using a jackknife resampling procedure. We reapplied the averaging-over-orderings procedure to the data set with one neuron left out and computed 95% confidence intervals over the distribution of all partial datasets.

t-SNE visualization

To visualize the structure of the data we used a method for dimensionality reduction known as t-distributed stochastic neighbor embedding (t-SNE) [233], a variant of the stochastic neighbor embedding technique originally developed by Hinton and Roweis [96]. This method attempts to minimize the divergence between the distributions of neighbor probability in the high-dimensional space and a low-dimensional space. The input to the algorithm was a set of 225 data vectors,

each of which collected the firing rates of all neurons in an area in response to a stimulus. We also normalized the data so that, for each neuron, responses to the 225 images had mean 0 and standard deviation 1. We used an initial dimensionality reduction down to 90 dimensions, and perplexity value of 30. We ran the algorithm multiple times to ensure convergence and stability of map estimates.

Classification decoding

We used a simple Poisson decoder to classify samples or families into one of 15 different categories. On each iteration we randomly selected a number of units from our recorded population. Since our units were recorded sequentially, we randomized the order of repetitions for each cell. To compute performance in the sample classification task, we estimated the mean spike counts of each neuron for each of the 15 samples within each family by computing the sample average over 10 of the 20 repetitions. For the held-out 10 repetitions of each sample, we computed which of the 15 samples was most likely to have produced the population response, assuming independent Poisson variability under the estimated mean spike counts. We computed the average performance (% correct) over all samples and families and repeated this process 10,000 times to get a performance for each population size. To compute performance in the family classification task, we estimated the average spike counts for each family over eight of the 15 different samples, and all repetitions. For each of the repetitions of the held-out seven samples we computed which of the 15 families was most likely to have produced the population response. We computed the average performance over all repetitions and repeated this process 10,000 times to get a performance for each population size. Results were similar using several alternative decoding methods including

a linear classifier and a mixture-of-Poissons model. The potential advantage of a more sophisticated mixture-of-Poissons model was negated by the larger parameter space and insufficiency of data. We also performed family classification by training on a subset of repetitions over all samples, and found increased performance in both V1 and V2, although V2 still outperformed V1.

Matched subpopulation decoding

To examine the effect of receptive field properties on decoding, we excluded neurons with a modulation ratio (f1/f0) greater than 0.8 and extracted 40-neuron subpopulations in each area that were matched for the mean and variance of classical receptive field (CRF) size (mean CRF in both V1 and V2 = 0.73±0.02°). We decoded our CRF-matched, complex cell subpopulations and compared performance to that achieved by 40-neuron subpopulations sampled randomly from the full population of both areas (mean CRF in V1 = 0.62±0.05°; V2 = 1.1±0.09°). In the sample classification task, V1 performance was significantly reduced by drawing matched subpopulations (65% to 55%), and there was no effect on V2 performance (which remained at 46%). V1 performed significantly better than V2 in sample classification for both unmatched ($p < 0.005$; bootstrap test resampling neurons and cross-validation partitioning) and matched subpopulations ($p < 0.01$). In the family task, V1 performance was increased by drawing matched subpopulations (30% to 35%), and V2 performance was only slightly decreased (41% to 40%). V2 performed significantly better than V1 in family classification for both unmatched ($p < 0.05$) and matched subpopulations ($p < 0.05$).

Discrimination decoding and prediction

We used the same decoding procedure for family classification, but performed discrimination between all pairs of texture families, yielding 105 pairwise comparisons. All discrimination decoding was performed using 100 units and was repeated 10,000 times to get a performance value. We transformed the measured performance values for V1 and V2 into a d' value and performed total least squares regression to get a linear fit to the V1 and V2 data. We then isolated two subsets of parameters from the texture model. The first consisted of the correlations of linear filter responses at nearby locations, which represent second-order pixel statistics, and are most intuitively described as representing a portion of the power spectrum (as such, we refer to them as "spectral"). We also gathered a set of higher-order statistics, consisting of correlations of magnitudes at neighboring locations, orientations, and scales, as well as correlations of phase-adjusted filter responses at adjacent scales [172].

To summarize the family discrimination capability of each group of statistics, we computed a matrix whose columns contained the absolute value of the difference between those statistics for each pair of texture families (105 columns, one for each pair of families). For the spectral statistics, we reduced the dimensionality (number of rows) of this matrix using principal components analysis. We found that four components captured 70% of the variance, and standard regression analysis revealed that both V1 and V2 performance was well predicted by a weighted sum of these components, (Fig. 3.7b). To examine the relationship between higher-order statistics and neural performance, we first removed the effects of the spectral statistics. We adjusted each of the rows of the higher-order differ-

ence matrix by projecting out the dimensions spanned by the rows of the spectral difference matrix. We then again reduced the dimensionality (number of rows) of this matrix using principal components analysis, retaining those components needed to capture at least 70% of the variance (in this case, ten components). Regression analysis revealed that a weighted sum of these components provided a good prediction for the difference in performance between V2 and V1 (Fig. 3.7c).

3.3 Examining tolerance of single neurons

We studied the population representation of visual information in areas V1 and V2 using naturalistic images generated from a texture model defined in terms of joint and marginal statistics of a simulated population of V1 simple and complex cells [172]. These statistics include local correlations between the output of pairs of model neurons that differ in preferred spatial frequency, position, and/or orientation. Some of these correlations are second-order statistics that capture the amount of energy at specific orientations and spatial frequencies — we refer to these as "spectral" statistics. Other correlations are higher-order, capturing naturalistic features beyond the power spectrum. We first computed this set of statistics for a grayscale photograph of a natural texture, and then generated synthetic texture images by starting with an image of Gaussian white noise and iteratively adjusting the pixels until the image had the same statistics (computed over the entire extent of the synthesized image) as the original photograph [172].

We refer to a set of images with identical statistics as a texture "family" (Fig. 3.1a, columns). Within a family, different white noise starting images yield different synthetic images, and we refer to all such images as "samples" from that family (Fig. 3.1a, rows). By construction, samples are identical in their model statistics,

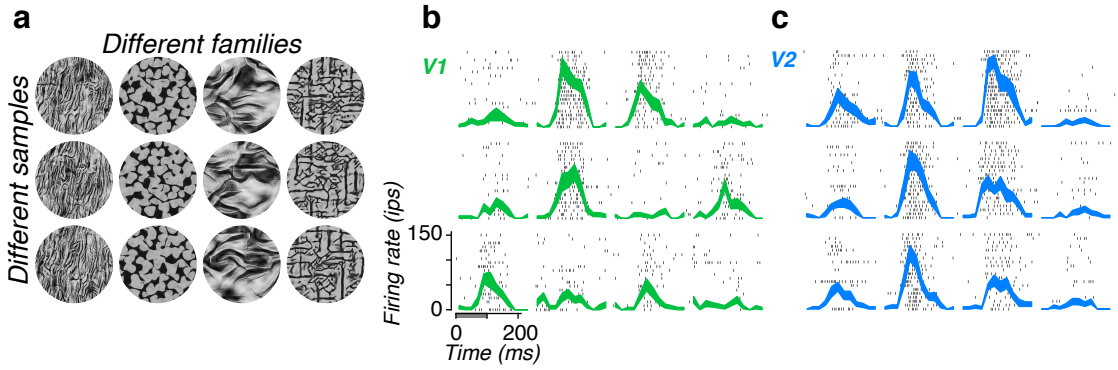

Figure 3.1: Example responses of V1 and V2 neurons. (a) Naturalistic texture stimuli. Each column contains three different samples from each of four texture families. The samples within each family are statistically matched, but differ in detail because the synthesis procedure is initialized with independent images of Gaussian white noise. (b) Raster plots and mean firing rates for an example V1 neuron, responding to textures in (a). Gray bar indicates presentation of the stimulus (first 100 ms), and each row of black ticks represents the timing of spikes on a single presentation of the stimulus. Thickness of lines indicates s.e.m across 20 repetitions of each of the images in (a). (c) Same as in (b), for an example V2 neuron.

but differ in the location and arrangement of features across the image. Previous work [172, 16] and visual inspection of Fig. 3.1a reveals that samples from a given family are similar in appearance to each other, and to the original photograph from which their statistics were drawn. We recently showed that these stimuli produce enhanced responses in V2 neurons, compared to images that are matched only for their Fourier power spectra [61]. This enhancement was not found in V1 neurons.

For the present study, we chose 15 original natural photographs to define 15 different texture families. These images were perceptually distinct and human sensitivity to their higher-order statistics spanned a range that was similar to that found over a much larger set of natural photographs [61]. We synthesized 15 different samples from each family, yielding 225 unique images.

We recorded the spiking activity of 102 V1 and 103 V2 neurons in 13 anesthetized macaque monkeys to these texture stimuli. We presented the stimuli

within a four-degree aperture centered on the receptive field of each recorded neuron. Each of the 225 different stimuli appeared 20 times in pseudorandom order, and was displayed for 100 ms, separated by 100 ms of uniform gray at the mean luminance. The same stimulus sequence was presented to each neuron. We have previously published a comparison of these responses to those obtained from spectrally matched (phase-scrambled) noise stimuli [61]. Here, we present a new analysis of these data that seeks to determine the relative selectivity and tolerance of V1 and V2 neurons for the different texture families and the image samples drawn from those families, respectively.

Texture stimuli elicited selective responses in most V1 and V2 neurons (Fig. 3.1b,c). Neurons in both V1 and V2 displayed a characteristic firing rate for each image, with some variability across presentations. For most texture families, firing rates of V1 neurons were highly variable across the samples (Fig. 3.1b). In contrast, V2 neurons exhibited similar firing rates across samples, as well as more consistent differences in average firing rate across families (Fig. 3.1c). That is, V2 neurons appeared to be more tolerant to the variations in image detail that occur across samples within a texture family, and more selective for the statistical parameters that define the family.

To quantify this observation, we used a nested analysis of variance (ANOVA) to partition the total variance in firing rate for each neuron into three components representing variation across families (columns), across samples within a family (rows), and across repeated presentations of each sample (residual spiking variability across rows of each raster in Fig. 3.1b,c). We first note that a smaller portion of V2 response variance was explained by the stimulus, as compared to V1 (Fig. 3.2a,b insets), consistent with previous reports [74]. The reduction in

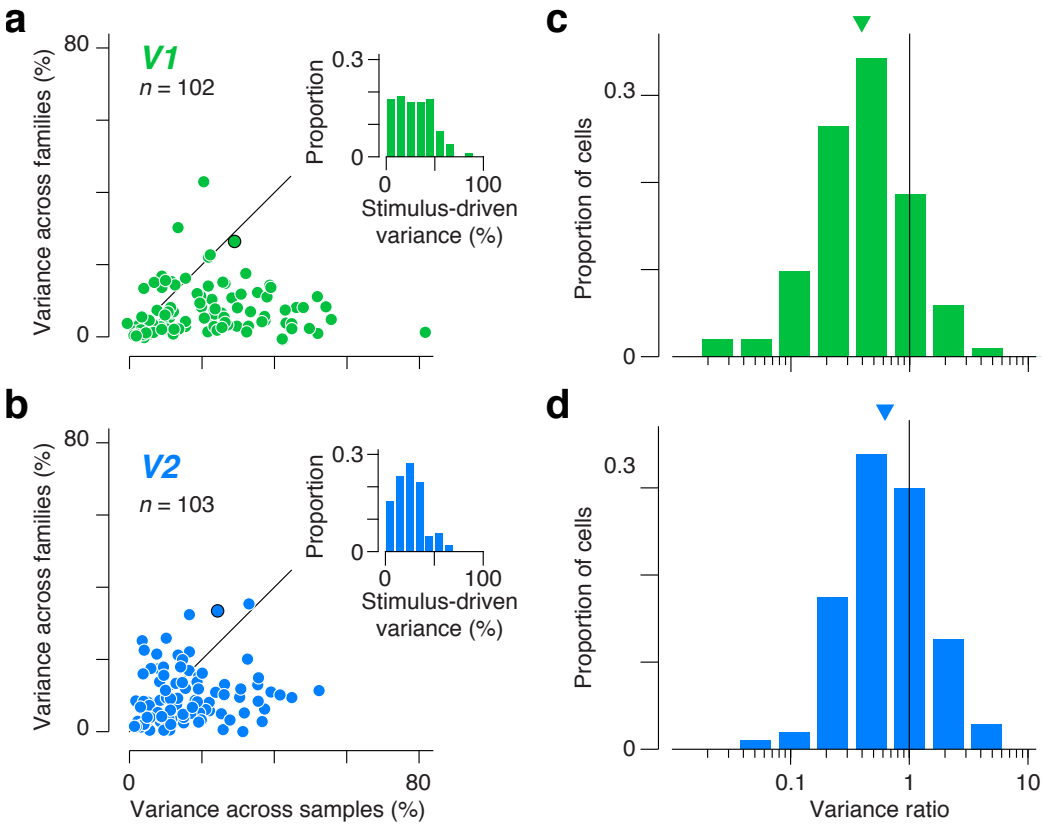

Figure 3.2: Nested ANOVA analysis of single-unit responses in V1 and V2. (a,b) Response variance of single units in V1 and V2 is partitioned into a component across families, a component within families (across samples), and a residual component across stimulus repetitions (noise). The position of each dot indicates, for a single neuron, the percentage of variance corresponding to the first two of these components. The insets indicate the distribution of the sum of these first two components. Points outlined in black correspond to the example single units shown in Figure 1. (c,d) Distributions of the ratio of across-family versus across-sample variance for V1 and V2. The geometric mean variance ratio was 0.4 in V1 and 0.63 in V2 (indicated by triangles). The difference was significant ($p < 0.001$, t-test in the log domain).

explainable variance in V2 was accompanied by a reduction in the population average firing rate compared with V1 (8.3 ips in V2 compared with 13.6 ips in V1), and may reflect a greater effect of anesthesia in V2.

While variance within families dominated the responses of most V1 neurons, many V2 neurons exhibited as much or more variance across families. However, the absolute levels of variance across and within families are affected by our particular choice of texture stimuli. To eliminate the influence of the stimulus ensemble, we compared the ratio of variance across and within families for neurons in V1 and V2 (Fig. 3.2c,d). This ratio is similar to the F-statistic from our ANOVA analysis, with a large value indicating high tolerance to the statistical variation of samples within families for our stimulus set. We found a significantly larger value of the variance ratio in our population of V2 neurons compared with V1 (Fig. 3.2c,d; $p < 0.001$, t-test on the log variance ratio). 29% of neurons in V2 were more variable in their firing rate across families versus within families compared with 16% of V1 neurons. These data indicate that on the whole, the V2 population exhibited more stable responses across samples within a family.

3.4 Analyzing the influence of receptive field properties on tolerance

We wondered whether this difference in tolerance was a consequence of well-known differences in receptive field properties between V1 and V2. For example, V2 contains a larger proportion of neurons that can be classified as complex (as opposed to simple [101, 123]) and the receptive fields of V2 neurons at a given eccentricity are about twice as large as those in V1 [64, 201]. Both of these properties

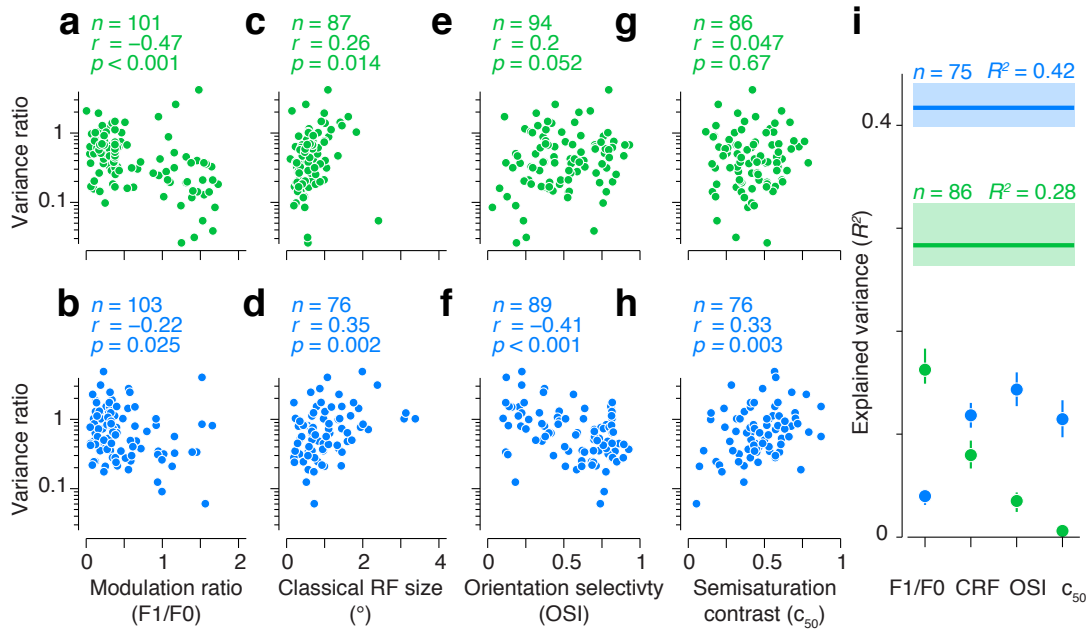

Figure 3.3: Dependence of response tolerance on conventional receptive field properties. (a–h) Variance ratio plotted against receptive field properties of individual V1 (green) and V2 (blue) neurons. (i) Results of multiple linear regression of the variance ratio against the four receptive field properties highlighted in (a–h). Horizontal lines show total explained variance for V1 (green) and V2 (blue). Points represent the contribution to explained variance from different receptive field properties determined using the averaging-over-orderings technique [77]. Shaded regions and error bars represent 95% confidence intervals computed using jackknife resampling.

would be expected to contribute to the variance ratio. Specifically, simple cells are sensitive to phase and should exhibit more response variation than complex cells across samples. Similarly, neurons with small receptive fields have a more limited area over which to compute statistics, and thus their responses are expected to fluctuate with changes in local statistics across samples (note that the statistics of sample images within a family are identical only when measured across the entire image).

To examine these and other effects on the variance ratio, we measured responses of a subset of our V1 and V2 populations to drifting sinusoidal gratings, and used

these to quantify ten conventional receptive field properties. We then used a stepwise regression separately in both areas to determine which of these properties might explain the within-to-across-family variance ratios (see methods). Altogether, receptive field properties accounted for only a limited amount of diversity of the variance ratios in both areas (Fig. 3.3i; V1, $R^2 = 0.28$; V2, $R^2 = 0.42$). This was not due to data insufficiency in our estimation of the variance ratio, since one half of our data could predict the other accurately (V1, $R^2 = 0.89 \pm 0.02$; V2, $R^2 = 0.86 \pm 0.02$; mean and standard deviation of bootstrapped distribution; see methods). As expected, we found that size and the spatial phase sensitivity of receptive fields were significantly correlated with the variance ratio, and this relationship held for both V1 and V2 (Fig. 3.3a–d). For V1 neurons, no other properties were significantly correlated (Fig. 3.3e,g). But in V2, the strength of orientation tuning (Fig. 3.3f) and contrast sensitivity (Fig. 3.3h) were also correlated with the variance ratio: Neurons with weaker orientation tuning and lower contrast sensitivity appeared to be more tolerant. To summarize these effects, we decomposed R^2 using the averaging-over-orderings technique [77] and examined the contribution of each property to the explained variance in V1 and V2 (Fig. 3.3i). This analysis confirmed the different pattern of contributions for the two areas.

We found that this pattern depended on the presence of higher-order correlations in the texture families. We computed the response tolerance for each neuron by taking the ratio of response variance across and within families that contained no higher-order statistics and only differed in their power spectrum (Fig. 3.4a, see chapter 2). We repeated the regression analysis on these variance ratios and found that, while the pattern of receptive field contributions remained the same

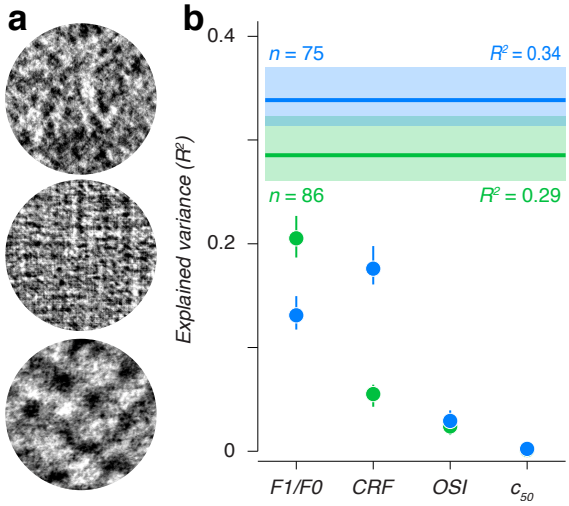

Figure 3.4: Response tolerance for spectrally matched noise stimuli. Results of multiple linear regression of the variance ratio measured from responses to spectrally matched noise stimuli. We assessed the contributions of the four receptive field properties highlighted in Fig. 3.3a–h. Format is the same as in Fig. 3.3i.

in V1, it differed markedly for V2 neurons (Fig. 3.4b). In particular the relationship between V2 tolerance with orientation selectivity and contrast sensitivity was eliminated. Instead, the pattern of contributions in V1 and V2 was very similar, with receptive field size and the modulation ratio explaining most of the variability in response tolerance. We conclude that while some of the increase in tolerance of V2 over V1 may be due to conventionally assessed differences in receptive field properties, some other factor is needed to fully explain the enhanced tolerance of V2 neurons to naturalistic stimuli.

3.5 Visualizing selectivity and tolerance of neuronal populations

We visualized the representation of texture stimuli within each neural population by transforming their responses from the high-dimensional response space (dimensionality = number of neurons) to a two-dimensional space. Ideally such a mapping would capture local and global aspects of the representation as much as possible. We used the t-distributed stochastic neighbor embedding (t-SNE) algo-

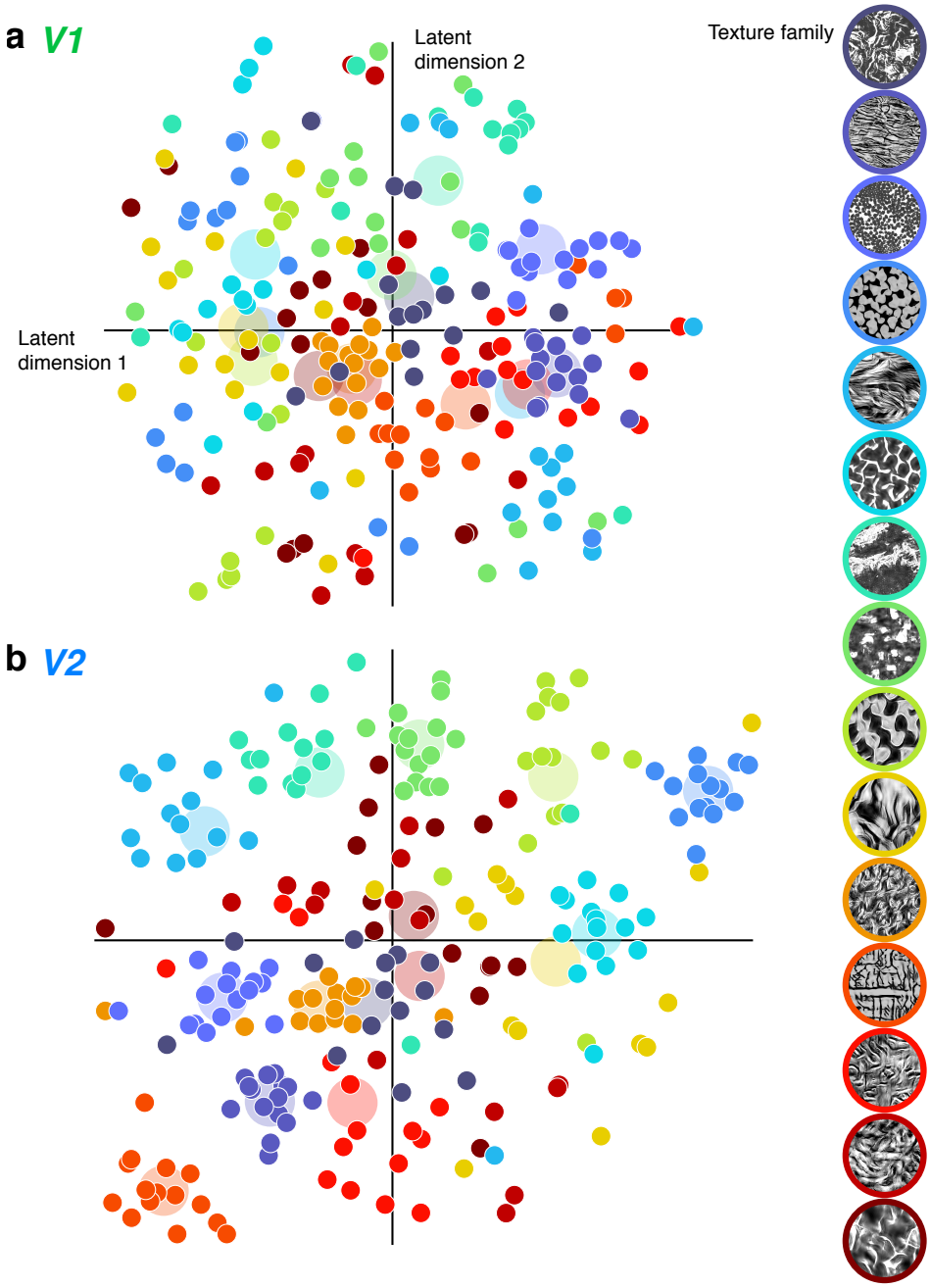

Figure 3.5: Two-dimensional visualization of neural population responses in V1 and V2. (a) V1 population response to each visual texture stimulus, displayed in a two-dimensional coordinate system that captures the full population of 102 V1 responses (computed using t-distributed stochastic neighbor embedding [233]). Each point represents one texture image, with color indicating the texture family. The larger, desaturated disks in the background indicate the centroid of all samples within each family. (b) Same analysis for the population of 103 V2 responses.

rithm, which solves this problem by minimizing the difference between the high- and low-dimensional distributions of neighbor distances [96, 233]. The choice of two dimensions is purely for interpretability and visualization, and is not meant not to imply anything about the underlying dimensionality of representation in either area.

We normalized the firing rate of each neuron and applied t-SNE to the V1 and V2 populations separately (Fig. 3.5a,b). Each of the 225 points represents population responses to a single texture sample, colored according to the family to which it belongs. Points that lie close together correspond to images that evoked similar responses from the neural population. Within V1, the groups of images from the same family generally produce scattered population responses, and the closest neighbors of most images do not correspond to samples from the same texture family (Fig. 3.5a). When applied to V2, the visualization reveals that population responses often cluster by texture family (Fig. 3.5b), with all the samples from several families tightly grouped.

3.6 Decoding neuronal populations

The low-dimensional t-SNE visualization (Fig. 3.5) provides an intuition for how the representation in V2 differs from V1, which can be more precisely quantified using a neural population decoder. To this end, we analyzed the ability of V1 and V2 representations to support two different perceptual discrimination tasks. For the first task, we built a Poisson maximum-likelihood decoder to discriminate between the 15 different samples within a texture family based on the responses within a neural population (Fig. 3.6a; see methods). Performance in both areas, averaged across all texture families, increased as the number of neurons included

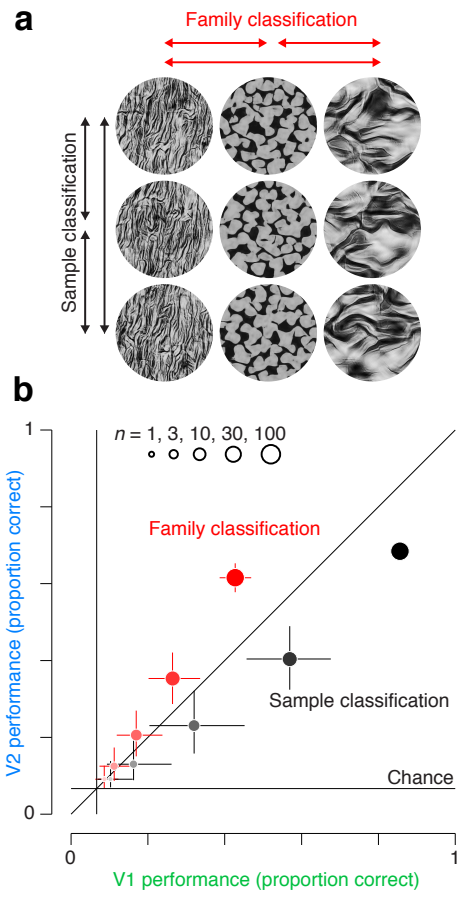

Figure 3.6: Quantifying representational differences between V1 and V2. (a) Schematic of sample (black) and family (red) classification. For sample classification, holdout data were classified among the 15 different samples for each family. Performance for each of the families was then averaged together to get total performance. For family classification, the decoder was trained on multiple samples within each family, and then used to classify held-out data into each of the 15 different families (b) Comparison of proportion correct classification of V1 and V2 populations, for family classification (red) and sample classification (black). We computed performance measures for both tasks using five different population sizes, indicated by the dot size. Chance performance for both tasks was 1/15. Error bars represent 95% confidence intervals of the bootstrapped distribution over included neurons and cross-validation partitioning.

in the analysis increased, but V1 outperformed V2 for all population sizes (Fig. 3.6b). The representation of image content in V1 thus provides more information for discriminating between specific samples.

For the second task, we built another decoder to discriminate between the 15 different texture families (Fig. 3.6a; see methods). We forced this decoder to generalize across samples by training on a subset of samples and testing on samples not used in the training. For both V1 and V2, and for all population sizes, absolute performance on this task was worse than on the sample classification task, although the difference was much larger in V1 (Fig. 3.6b), but in contrast to the sample classification task, V2 outperformed V1 for all population sizes. To

examine whether this finding could be a consequence of the differences in receptive field properties described above (Fig. 3.3), we excluded neurons classified as simple from both areas and selected subpopulations matched for classical receptive field size. This matching procedure had little effect on V2 performance in either task, but did reduce V1 performance on the sample task and increase V1 performance on the family task (see methods). However, performance in the two areas remained significantly different, suggesting more complex forms of selectivity are involved.

3.7 Comparing selectivity of neuronal populations

To elucidate the V2 response properties that allow it to outperform V1 in family classification, we examined the dependence of performance on the differences in statistics between pairs of texture families. We built a Poisson maximum-likelihood decoder to best discriminate between each pair of texture families (105 different comparisons). Comparing performance in V1 and V2 (Fig. 3.7) reveals two promi-

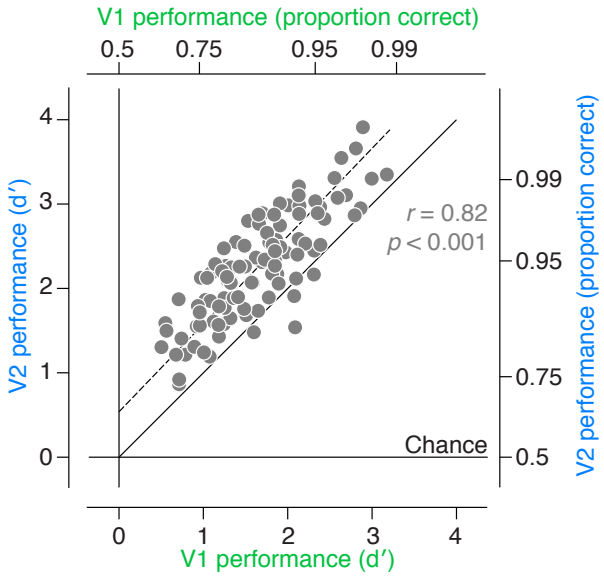

Figure 3.7: Texture discrimination performance of neural populations. Comparison of V1 and V2 performance on pairwise texture discrimination. Performance values were plotted on coordinates that varied linearly in discriminability (d′). Right and top axes indicate corresponding values of performance expressed as proportion correct. Each point represents one of 105 pairwise comparisons among the 15 texture families. Solid lines indicate chance performance and the identity line. Dashed line indicates best fit using total least squares.

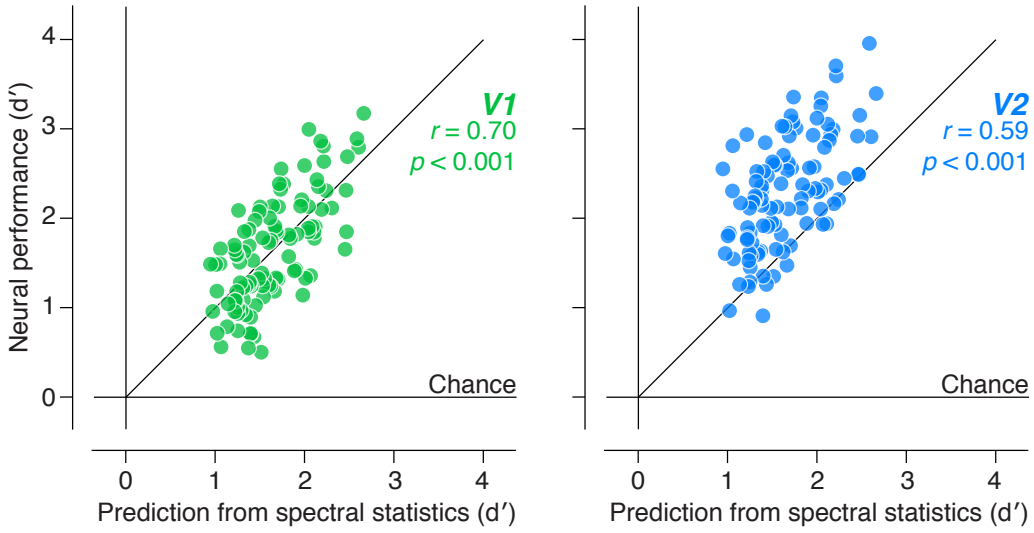

Figure 3.8: Comparison of V1 and V2 performance to that of a model capturing spectral statistics. The magnitude of difference in spectral statistics for each texture family pair was weighted to best account for the performance of V1. Both V1 (left) and V2 (right) performance were plotted against this spectral prediction.

nent features (Fig. 3.7). First, performance in V1 and V2 was highly correlated across the different texture discriminations ($r = 0.82$; $p < 0.001$), suggesting that some of the features that drive performance in V1 are also responsible for performance in V2. Second, V2 neurons performed better for nearly all pairs, and this improvement was approximately independent of the performance seen in V1 (Fig. 3.7). A straight-line fit suggests that if V1 discrimination performance were at chance, V2 performance would be 65% correct ($d' = 0.54$). To understand this relationship, we sought to separate those stimulus properties that drive performance in both V1 and V2 from those that underlie the increase in performance of V2 over V1.

We chose texture families for this study that differed in their spectral content — the relative amount of energy at different orientations and spatial frequencies. V1 neurons are highly selective for spectral content [42] and this selectivity is main-

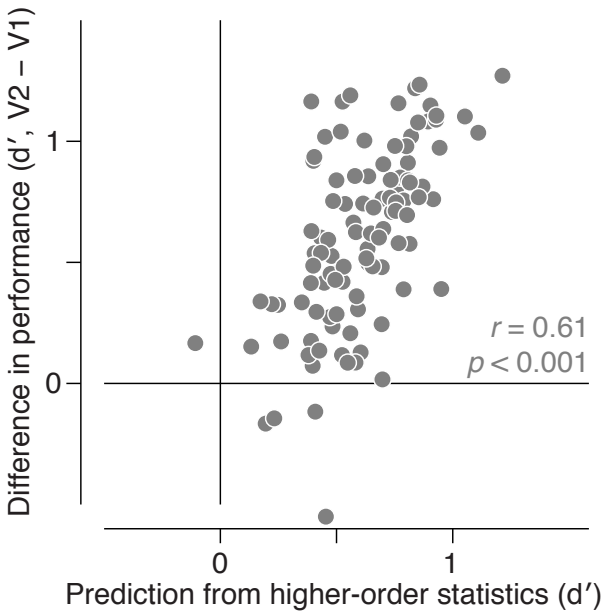

Figure 3.9: Discrimination performance predicted by higher order statistics. Comparison of the difference in V1 and V2 performance to the strength of higher-order correlation differences. The magnitude of difference in higher-order correlations for each texture family pair was weighted to best predict the difference in V1 and V2 performance. Performance was first transformed into discriminability (d′) units.

tained in V2 [123]. We wondered whether the spectral parameters of the model could explain V1 performance. Across all 105 pairs of texture families, we measured the magnitude of the difference in spectral statistics between the two families. We then predicted V1 discrimination performance from the statistical differences, over all pairs (Fig. 3.8; see methods). The spectral differences predicted V1 performance well ($r = 0.7$; $p < 0.001$), and the same model also provided a good prediction for V2 performance ($r = 0.59$; $p < 0.001$). Reoptimizing the weights to predict V2 responses barely improved the correlation ($r = 0.6$; $p < 0.001$), consistent with the notion that the spectral information represented in V2 is directly inherited from V1. However, the spectral statistics captured little of the difference in performance between V1 and V2 ($r = 0.22$; $p < 0.05$).

These analyses suggest that the superior performance of V2 must be due to the higher-order (i.e., beyond second-order) correlations present in the texture model. To test this hypothesis, we extracted the parameters that capture higher-

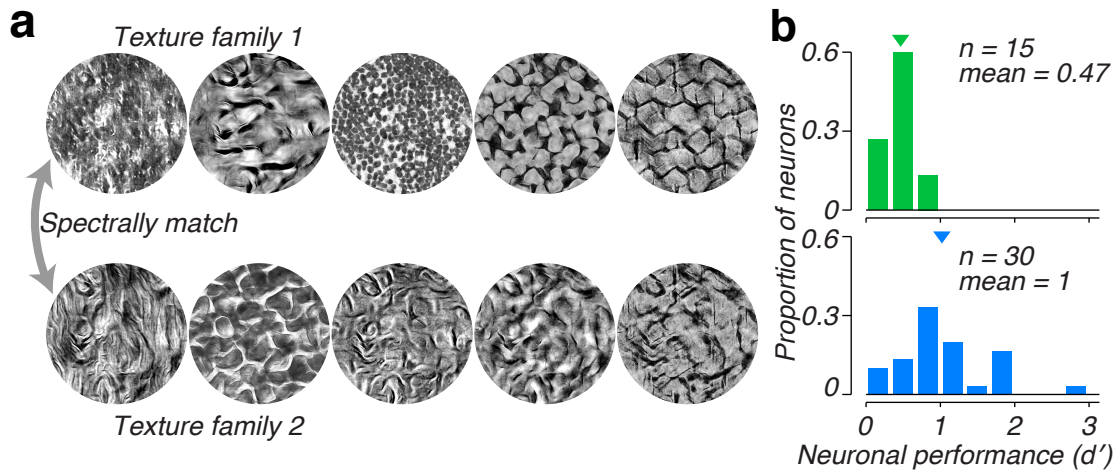

Figure 3.10: Single neuron discrimination between spectrally matched families. (a) Examples from five pairs of texture families that have been spectrally matched. For each column, the top and bottom texture families differ only in their higher-order statistics. (b) Distribution of maximum discriminability across single neurons in V1 and V2. For each neuron, we performed an ideal observer analysis on the spike counts to multiple samples of the five pairs of spectrally matched texture families shown in (a), and selected the texture family pair with maximum d′.

order statistics through correlations of filter response magnitudes across position, frequency, and orientation, and projected out the portion captured by the spectral statistics. We then predicted the difference in V1 and V2 performance (Fig. 3.9). Differences in the higher-order statistics — in contrast to spectral statistics — provided a good prediction for the V1/V2 performance difference ($r = 0.61$; $p < 0.001$).

To provide a more direct test and confirm this prediction, we subsequently performed a tuning experiment on a small subset of V1 and V2 neurons. We created pairs of texture families that differed only in their higher-order statistics (Fig. 3.10a, see methods), and presented them to single neurons in V1 and V2 of an awake, fixating macaque. Here we did not perform a population decoding analysis, but simply examined the maximum single-neuron performance at discriminating

higher-order statistics (Fig. 3.10b, see methods). We found several neurons in V2 that were highly sensitive to differences in higher-order statistics, and very few in V1. On average, single-neuron performance in V2 was over twice as high as in V1 (Fig. 3.10b).

In summary, V1 discrimination performance was well captured by the spectral statistics of naturalistic textures. This same set of statistics captured a significant portion of V2 discrimination performance, but most of the superiority of V2 over V1 comes from higher-order statistics.

3.8 Discussion

Our results support the hypothesis that populations of V2 neurons represent the statistics of the activity of local ensembles of V1 neurons, which capture the appearance of naturally occurring textures. Using a set of stimuli for which these statistics are tightly controlled, we showed that, relative to neurons in V1, V2 neurons exhibit increased selectivity for these statistics, accompanied by an increased tolerance for randomized image variations that do not affect these statistics. This "tolerance to statistical resampling" complements the more widely discussed visual invariances to geometric distortions (e.g., translation, rotation, dilation) [263, 184], or changes in the intensity, color, or position of a light source [242, 12].

Our results also help to integrate and interpret other findings. The selectivity of V2 neurons for many artificial stimuli, including gratings, angles, curves, anomalous contours, and texture-defined patterns, is nearly the same as that of V1 neurons [210, 89, 91, 120, 132, 106, 11, 196]. This similarity would be expected if V2 neurons are selective for a broad set of V1 response statistics, and not for a small subset of specialized combinations of V1 inputs as assumed by

these approaches. On the other hand, the tolerance of V2 cells identified here does seem consistent with the previously identified behaviors of "complex unoriented" V2 cells [99], which are selective for patches of light of a particular size but tolerant to changes in position over a much larger region. Such a property may explain why orientation selectivity so strongly predicted tolerance in V2 but less so in V1. This relationship might also reflect greater heterogeneity of orientation tuning within V2 receptive fields [11] providing a substrate for computing local orientation statistics.

Our results complement recent work demonstrating V2 selectivity for third- and fourth-order pixel statistics. Yu and colleagues [256] examined responses of V1 and V2 neurons to binary images synthesized with controlled pixel statistics up to fourth order, and found that neuronal selectivity for multipoint (i.e., third and fourth order) correlations is infrequent in V1, but common in V2. The strength of this work derives from the well-defined stimulus ensemble, which covers the full set of statistics up to fourth order and allows a thorough assessment of the selectivity for individual statistics in the responses of single neurons. On the other hand, the restriction to statistics of a particular order, while mathematically natural, is not necessarily aligned with the restrictions imposed by the computational capabilities of biological visual systems, and this may explain why selectivity of V2 neurons for these statistics is only modestly greater than that of V1 neurons. The stimuli in our experiments are constrained by statistics that are defined in terms of an idealized response model for a V1 population. Although they also constrain multipoint pixel statistics, they do not isolate them in pure form, and they span too large a space to allow a thorough experimental characterization of selectivity in individual cells. On the other hand, they represent quantities that may be more directly related to

the construction of V2 responses from V1 afferents, and they allow direct synthesis of stimuli bearing strong perceptual resemblance to their ecological counterparts [172, 16, 17, 60].

The particular statistics we matched to create our texture families are surely not represented fully and only in V2, and this may explain why the reported difference in selectivity and tolerance between V1 and V2, while robust, is not qualitative. In particular, these statistics include both the local correlation of oriented linear filter responses (equivalent to a partial representation of average spectral power across the image) as well as pairwise correlations between the magnitudes of responses of oriented filters tuned to different orientations, spatial frequencies, and positions. We created different families from the statistics extracted from 15 original photographs, which differed in both the spectral and higher-order statistics. We found that the spectral differences between different families accounted for a substantial portion of the discrimination performance of both V1 and V2 populations. However, V2 nearly always outperformed V1, and this superiority was well predicted by the differences in higher-order statistics. This suggests that an artificial stimulus set in which families differ only in higher-order statistics would better differentiate the discrimination performance of V1 and V2.

How do V2 neurons achieve higher classification and discrimination performance than their V1 inputs? There are two possible answers: reducing variability in the representation of individual families or increasing the mean separation in the representations of different families. The first of these can be achieved by combining many V1 inputs so as to average out their trial-by-trial variability. Larger receptive fields may be an indication of such a construction: Indeed, larger receptive fields are associated with higher variance ratios and better family classification

performance. However, when we matched receptive field sizes between the two areas, V2 still performed better in family classification. Thus, we posit that V2 neurons are also taking advantage of the second option, transforming their V1 inputs to make family differences more explicit in their average responses. This transformation amounts to "untangling" the representation of visual features that were not directly decodable in the V1 representation [45]. Specifically, V1 neurons don't appear to signal the presence of higher-order correlations with a consistent change in firing rate, while V2 neurons do [61, 256]. A a result, larger differences in higher-order correlations between families explains a significant portion of the increased discrimination performance in V2 (Fig. 3.7c).

Perceptually, invariances related to statistical resampling were originally proposed by Julesz as a testable prediction of statistical texture models [108], and have been used to test and refine such models in both vision [172, 16, 222, 235] and audition [137]. Theories regarding the statistical summary nature of "crowding" effects in peripheral vision [122, 166, 167, 76] have also been tested for such perceptual invariances [17, 60], and are consistent with representation of texture statistics in area V2. Although our analysis of V2 responses is qualitatively consistent with these perceptual observations, the connection is difficult to test quantitatively. In particular, the statistics in our texture stimuli were computed by averaging over the full stimulus aperture, which was held to a fixed size of four degrees for all cells to allow a reasonable interpretation of population decoding. This size was generally larger than the receptive fields of the neurons (Fig. 3c,d). Thus, most neurons saw only a portion of the stimuli, over which the statistics would not have been fully matched. Finally, recall that the transformation from V1 to V2 is part of a cascade, and it may well be that perception relies more on downstream areas

such as V4 where neurons may show even more selectivity and tolerance for the statistics we used [162, 12].

The visual world is often described in terms of forms or "things" made up of lines, edges, contours, and boundaries, and these symbolic descriptions have played a dominant role in developing theories for both biological and machine representations of visual information. But textures and "stuff" [5] are ubiquitous in natural visual scenes, and are not easily captured with edge or contour descriptions. The results presented here suggest that V2 neurons combine V1 afferents to represent perceptually relevant statistical features of visual textures. It is currently unknown whether this statistical description of the visual world is also sufficient to account for perception of visual form. Recent work suggests that textural statistics such as those used here can account for aspects of peripheral vision that are not exclusive to the perception of texture [17, 60]. Additionally, recent successes in machine recognition of complex objects using multistage neural networks also call into question the need for explicit boundary, contour, or edge information in high-level vision. Indeed, the model responses at different stages of these neural networks have provided a good basis for accounting for neural responses in both midlevel and late stages of visual cortex [254, 112], and attempts to visualize the selectivities of model units at midlevel stages have often revealed texture-like visual structure [258]. Thus, the two-stage representation we describe here may provide a foundation for the representation of the more complex and structured signals found in images of objects or of entire visual scenes [122].

Chapter 4

Sensitivity from beyond the receptive field in V2

4.1 Introduction

We previously found that V2 neurons with larger receptive fields were no more likely to be sensitive to the higher-order statistics of natural textures than those with smaller receptive fields (Fig. 2.9a). However, we also observed that neuronal sensitivity in V2 was decreased by shrinking the diameter of the stimulus to match that of the estimated receptive field, suggesting that the surround may play a role in establishing sensitivity to naturalistic image structure in V2 (Fig. 2.9b). The influence of stimuli presented outside the classical receptive field has been studied extensively in V1 and is generally suppressive [9, 21, 43, 124, 34]. The strength of surround suppression is modified by several stimulus properties, including orientation, spatial frequency, and spatial position [35]. In general, the influence of the surround has not been posited to play a significant role in establishing or enhancing selectivity for visual features more complex than those represented within the receptive field center (but see [71, 83]). More often, surround mechanisms have

been proposed to increase the transmission efficiency for information presented to the receptive field center [193]. This efficiency hypothesis has mostly been tested by using natural images rather than the simpler artificial stimuli frequently used to characterize V1 neurons [236, 237, 38]. Surround selectivity has been less extensively studied in V2 but is generally considered to function similarly to that in V1 [201]. However, stimuli containing naturalistic structure have rarely been used to study the surround of V2 neurons, and it is currently unknown whether the surround in V2 may play a distinctive role in establishing selectivity for naturalistic visual features.

Here, we performed aperture-size tuning experiments on a population of V2 neurons using naturalistic textures and spectrally matched noise lacking higher-order features. Generally, responses were suppressed when both types of stimuli extended beyond the classical receptive field. However, responses in V2 also became increasingly sensitive to the higher-order structure of naturalistic textures as stimulus size increased. Part of this enhanced sensitivty was driven by increased surround suppression for spectrally matched noise stimuli. The dynamics of enhanced naturalistic sensitivity also mirrored the temporal onset of surround suppression to both types of stimuli. We conclude that surround suppression may be weakened by the presence of naturalistic correlations in the V1 input to V2 (an observation that runs opposite to recent findings on the effect of surround stimulation on V1 neurons [38]). Alternatively, a facilitatory mechanism selective for joint activity patterns of V1 neurons influences the firing rate of V2 neurons over a much larger region than the traditionally measured receptive field.

4.2 Methods

Physiological methods for this chapter were identical to those reported for chapter 2 in section 2.2.2.

4.2.1 Visual stimuli

We generated stimuli as described previously in section 2.2.1, with one exception. Using the Portilla-Simoncelli texture model [172], we analyzed the statistics of the 31 original photographs used in chapter 2. Each of these photographs had a resolution of 320 × 320. During synthesis, we began with an image of Gaussian white noise at a resolution of 2048 × 512, resulting in synthetic images at this larger resolution. When using these larger images in our experiments, we presented them at the same scale (pixels/degree) as our original experiments. For the initial characterization, we presented naturalistic and spectrally matched noise stimuli from each of these 31 families within a 4° raised cosine aperture, as in our initial experiments described in chapter 2.

Here, we showed different "samples" by shifting the center of this aperture horizontally across the 2048 × 512 image, rather than presenting multiple images that had been independently synthesized. These two methods are roughly consistent since the textures are spatially homogeneous and the statistics were analyzed for images much smaller than the 2048 × 512. However, the statistics of different "samples" here will be slightly less well matched than in the previous experiment since the statistics have only converged to match the original exactly when averaged over the full 2048 × 512 resolution. We analyzed the stability of the statistics when cropping from these larger texture stimuli and show that they have mostly

converged at the relevant size (Fig. 4.4a). When performing the size tuning experiments, we changed the aperture diameter applied to these larger textures. We always presented textures at the same scale of 80 pixels/° as in our intial experiments in chapter 2. For aperture sizes that exceeded 512 pixels (or 6.4°) the images were wrapped so the same part of the image was visible in two places on the monitor. This property likely had no effect on the results because of the large spatial segregation of the image elements and the homogeneous nature of texture stimuli.

4.2.2 Analysis

Receptive field characterization

We first performed size tuning experiments using drifting gratings presented for 1 s to the receptive field of each neuron. We calculated the response as the firing rate within a time window matched to the duration of the stimulus, shifted by the latency of each neuron. We fit a ratio of Gaussians model as described in [34] to the responses of each neuron. We then took twice the standard deviation of the excitatory Gaussian to be our estimate of the classical receptive field (CRF) size. Results were similar when using a different metric such as the optimal size. We then binned the stimulus diameters from each tuning curve by the estimated CRF size to get a relative size measure. Relative values between 0.1 and 0.4 were assigned to the 1/3 times CRF size bin, between 0.8 and 1.2 to the CRF-matched bin, and between 2.4 to 3.9 to the 3 times CRF size bin. We computed the surround suppression index from the maximum mean firing rate and the mean firing rate to the largest stimulus we presented. We subtracted the suppressed response from the maximum and divided by the maximum to obtain the fractional reduction.

We estimated the latency of each neuron by maximizing the stimulus associated response variance [212]. We aligned each neuron to its response onset and binned spikes at 1 ms, smoothing the peristimulus time histogram (PSTH) with a boxcar filter with a width of 10 ms. The time course of suppression and modulation indices were computed separately for each neuron and then averaged together.

Parameter stability

To examine parameter stability with stimulus size we first measured parameters of the Portilla Simoncelli texture model [172] from 15 samples of naturalistic and spectrally matched noise stimuli drawn from all 31 texture families used in our initial neural characterization. We separated out parameter groups as we have done previously (Fig. 2.22, [61]). Here, we focused on energy filter correlations across scale, position, and orientation because we showed them to be most important for establishing sensitivity to naturalistic visual structure (Fig. 2.22). We also used linear filter correlations across position because these capture the spectral sensitivity of V1 and V2 neurons well (Fig. 3.8), and should not be a cue for discriminating naturalistic from spectrally matched noise stimuli. We measured each individual parameter within a group across 15 samples of the 31 texture families used in our initial characterization. Each sample was measured at multiple image sizes by cropping the full image down to dimensions ranging from 64 $\times$ 64 pixels to 704 $\times$ 704 pixels (equivalent to 0.8° –8.8° of visual angle). For each texture family and image size we subtracted the parameter value (averaged over samples) to spectrally matched noise from that measured to naturalistic stimuli and divided by the sum of these two values, analogous to the computation of the modulation index we used for physiology. We averaged this parameter modulation

index across the 31 texture families and all parameters within a group to get our final measure.

4.3 Charactizing size dependence of naturalistic sensitivity

We generated synthetic stimuli containing naturally occurring marginal and joint statistics across the outputs of a simulated population of V1 simple and complex cells [172]. These statistics include spatially averaged local correlations between the output of pairs of model neurons that differ in preferred spatial frequency, position, and orientation. We first measured this set of statistics from a grayscale photograph of natural texture, and then generated synthetic texture images by starting with an image of Gaussian white noise and iteratively adjusting the pixels until the image had the same statistics as the original photograph [172]. We refer to images generated using the full set of parameters as "naturalistic" (Fig. 4.1a). We additionally created spectrally matched noise stimuli by randomizing the phase of each naturalistic image (Fig. 4.1b). These noise stimuli are matched to the originals in their average orientation and spatial frequency content, but

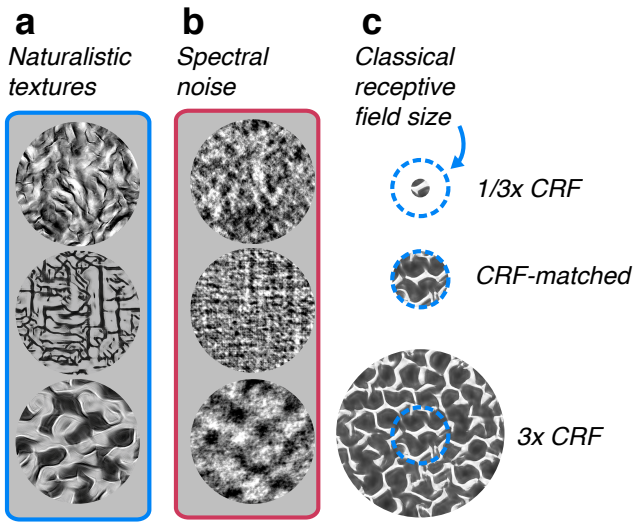

Figure 4.1: (a) Three examples of naturalistic textures. (b) Three examples of spectrally matched noise stimuli. (c) Design of the size tuning experiment. The receptive field size of an example V2 neuron is represented as a dashed blue line. Stimuli were centered on the receptive field and shown at a variety of sizes.

contain no higher-order structure. These two image types evoke equal firing rates in V1 neurons on average, while V2 neurons are driven more strongly by naturalistic stimuli [61].

We recorded neural responses from 42 V2 neurons in four anesthetized macaque monkeys to sequences of naturalistic and spectrally matched noise stimuli. For each neuron, we first characterized its classical receptive field (CRF) using drifting gratings to determine location and size. We subsequently measured responses to naturalistic and spectrally matched noise images presented within a 4° aperture centered on the receptive field [61]. We presented 15 "samples" of both naturalistic and noise stimuli drawn from 31 different texture "families" synthesized from 31 original images (one naturalistic and one spectrally matched noise sample from three families are shown in Fig. 4.1a and Fig. 4.1b, respectively). Each image was presented for 100 ms followed by 100 ms of mean luminance. We computed a modulation index from the responses to each texture family by subtracting the firing rates to naturalistic and spectrally matched noise and dividing by their sum. After this initial characterization, we chose a number of texture families for additional characterization based on the strength of this modulation index. For each chosen texture family we performed a size tuning experiment by varying the diameter of the aperture of the texture patch in logarithmically spaced intervals centered around our estimate of the neuron's classical receptive field size from its response to drifting gratings (Fig. 4.1c). Importantly, we varied the size of our stimuli by masking the full texture image, and not by rescaling the textures. The image content in the center of the receptive field was therefore identical for large and small size conditions (Fig. 4.1c).

We found a wide range of receptive field sizes and size tuning shapes across

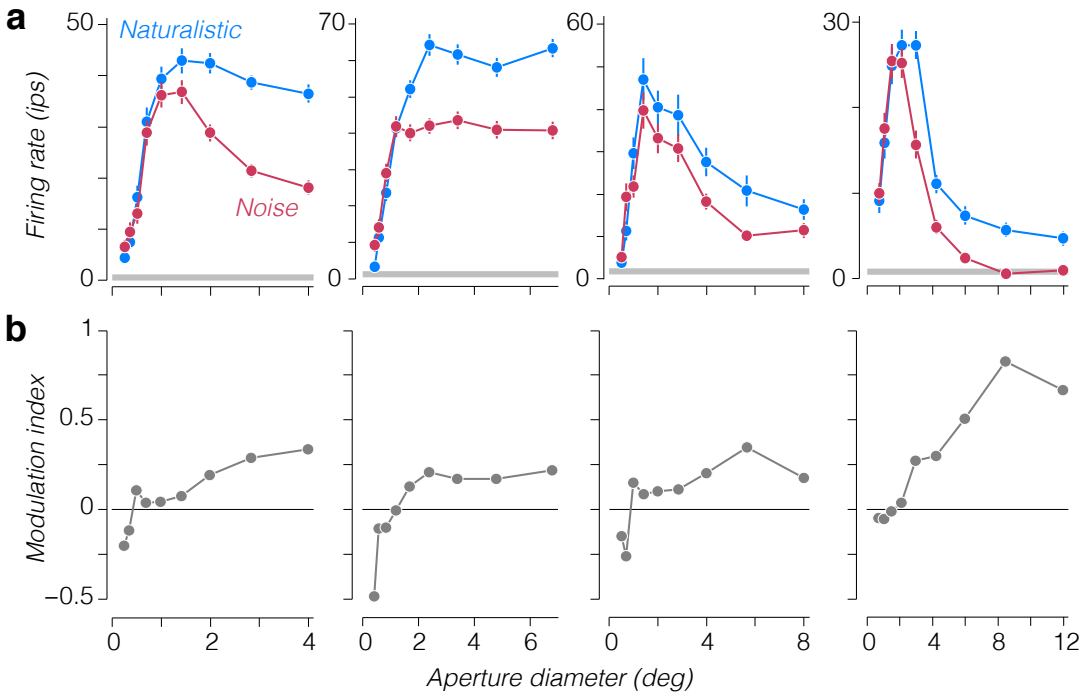

Figure 4.2: Size dependence of naturalistic sensitivity. (a) Firing rate to naturalistic and noise stimuli for four example V2 neurons as a function of the aperture diameter of texture patch. (b) Modulation index computed from the firing rates in (a).

our recorded population (Fig. 4.2a). Similar to our previous findings, we found no evidence for a relationship between the modulation index and estimated receptive field size using responses measured with drifting gratings ($r = 0.09$; $p = 0.55$). However, there were clear differences between size tuning to naturalistic and noise stimuli for most neurons. Specifically, naturalistic and spectrally matched noise stimuli appeared to drive similar responses at small sizes, while a preference for naturalistic textures emerged gradually as the size of the stimulus increased. We quantified this by computing a modulation index for each size (Fig. 4.2b). Modulation strength increased with aperture diameter for most neurons.

To examine this trend across the population we aligned each modulation tun-

ing curve to the estimated receptive field size of each neuron. We computed the average modulation strength across the population at three sizes: when the stimulus aperture was one-third the CRF diameter, roughly matched to the CRF, and three times the CRF (Fig. 4.3). We found that at sizes approximately one-third the CRF there was very little modulation across the population. Modulation was strong when aperture size was matched to the CRF, but roughly doubled when aperture size was increased to approximately three times the CRF size (Fig. 4.3). As with our previous results [61], this pattern suggested a role for the surround in strengthening sensitivity to naturalistic image structure.

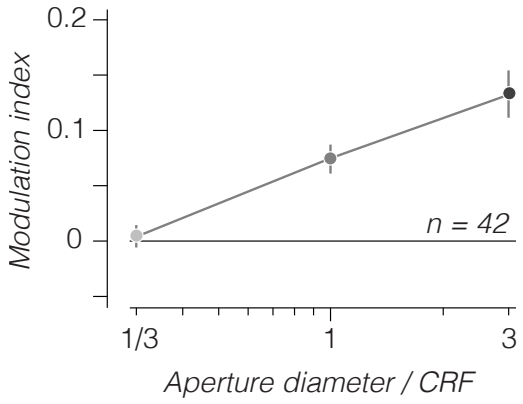

Figure 4.3: Population sensitivity depends on size. The average modulation index across the population was plotted against the relative stimulus size. Error bars represent ±s.e.m. across neurons.

A possible explanation for increasing modulation index with aperture diameter could be the increasing sample size for computing higher-order statistics. Our synthesis procedure only guarantees statistics to be fully converged to their specified values when averaging over the entire image. When the aperture is very small, the sample of higher-order statistics is potentially more variable and could lead to reduced sensitivity. To examine the influence of parameter stability on neuronal responses, we analyzed the spectral and higher-order statistics of samples of texture stimuli cropped at varying sizes. For different groups of parameters, we computed a modulation index, analogous to that used for neuronal responses, by taking the

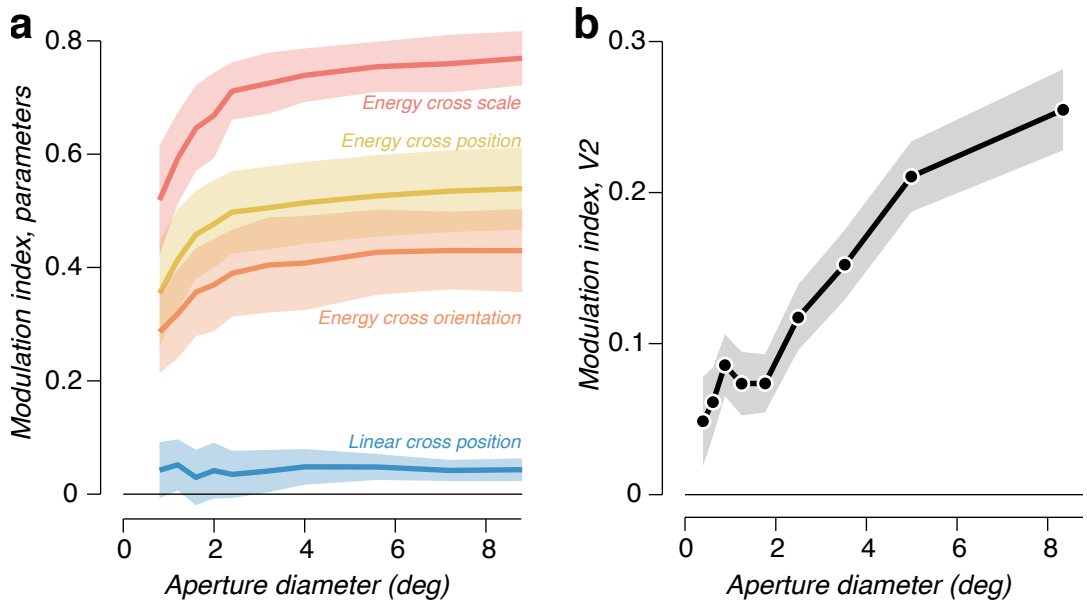

Figure 4.4: Parameter stability does not account for size tuning. (a) Modulation index computed on different groups of parameters between naturalistic and spectrally matched noise stimuli cropped to different sizes. Parameter groups are colored as in Fig. 2.22. Shaded regions indicate ±s.e.m. across individual parameters within a group. (b) Average modulation index from V2 neuronal responses as a function of stimulus aperture. The receptive field size of individual neurons is not taken into account. Shaded region indicates ±s.e.m. across neurons.

.

difference of each parameter value (averaged across samples) for naturalistic and spectrally matched noise stimuli and dividing by the sum (Fig. 4.4a). The spectral (second-order) statistics computed through correlations across the output of linear filters differing in position had a very small modulation index for all image sizes (Fig. 4.4a, blue). Although the stimuli are spectrally matched for large sizes, the variance of the spectral modulation index does increase at small sizes, which could drive neuronal response variability and diminish the modulation index.

We also examined parameter groups capturing magnitude correlations across orientation, scale, and position. The strength of the modulation index for all of these parameters was biased toward lower values when the image was small

(Fig. 4.4a). However, all groups appeared to approach their asymptotic values once stimuli reached about 2°, close to the average receptive field size of our V2 population (1.7°±0.8°). This contrasted with the average neuronal modulation index computed for different sized stimuli (Fig. 4.4b). The V2 modulation index more than doubled between 2° and the largest stimulus sizes around 8°. Therefore, instability in the higher-order statistics could potentially be driving the increase in modulation at small sizes, but could not explain the strong increase in sensitivity to naturalistic image structure we observed for texture patches extending well beyond the receptive field.

We wondered what could drive this enhancement of sensitivity as stimuli increased outside the CRF. Typically, the effect of contrast outside the receptive field has been found to be suppressive [9, 21, 43, 124, 34]. We observed a wide range of surround suppression for both naturalistic and noise stimuli, from no suppression (Fig. 4.2a, second from left) to nearly complete suppression (Fig. 4.2a, right). Although we previously found that surround suppression strength measured with drifting grating stimuli was not significantly correlated with the modulation index across neurons [61], we still wondered whether it might play a role. For each neuron, we computed a suppression index measuring the fractional reduction in firing rate to large stimuli [34]. There was a significant relationship between surround suppression strength measured with noise stimuli and the modulation index ($r = 48$; $p = 0.001$). However, it also appeared that modulation could be quite strong even in cases of little or no surround suppression (Fig. 4.2a, second from left). When we compared the strength of surround suppression to naturalistic versus spectrally matched noise stimuli, we found that there was significantly more suppression to noise stimuli across the population (Fig. 4.5a). We recorded a small

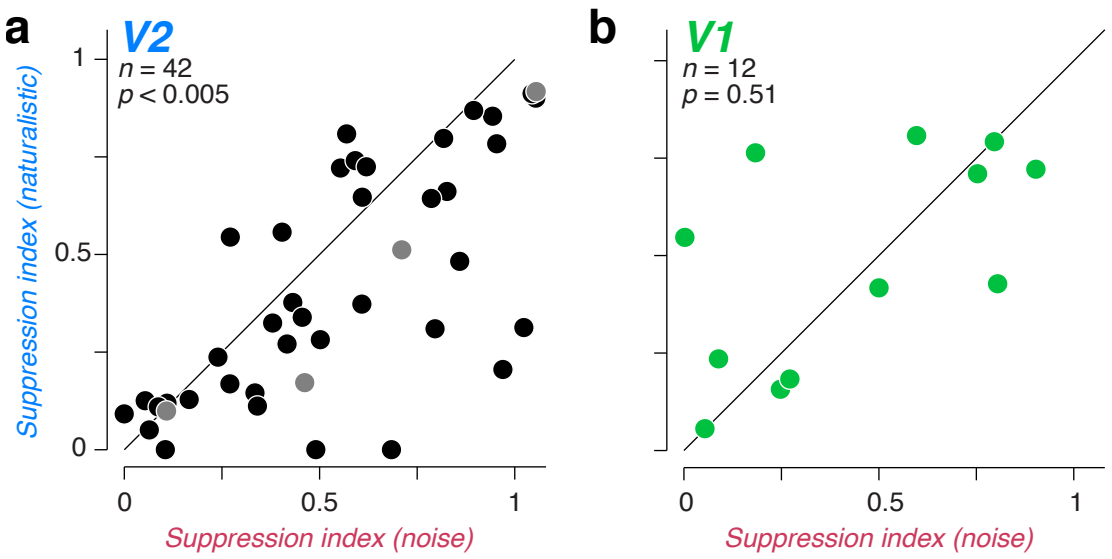

Figure 4.5: Surround suppression depends on naturalness in V2. (a) Surround suppression index measured form responses to spectrally matched noise compared with that measured to naturalistic stimuli. Each symbol represents a neuron, Suppression indices were averaged for neurons for which we collected multiple size tuning curves. Gray symbols represent the four V2 neurons shown in Fig. 4.2. (b) Same format as in (a) but for 12 V1 neurons.

population of V1 neurons under the same conditions and found no such effect (Fig. 4.5b), consistent with previous reports that used phase-randomized stimuli to study surround suppression in V1 neurons [38]. We conclude that areas outside the receptive field may play a distinctive role in V2 by enhancing sensitivity to naturalistic image structure.

To further examine the possible relationship between surround suppression and the strength of naturalistic modulation, we examined the time course of responses for these two effects. We aligned all our recorded neurons to their response onset and computed the average firing rate as a function of time for different conditions (Fig. 4.6a). We focused on trials where the stimulus diameter was roughly matched to the CRF of the neuron and when the stimulus was roughly three times the CRF diameter. We examined the time course of suppression between these two condi-

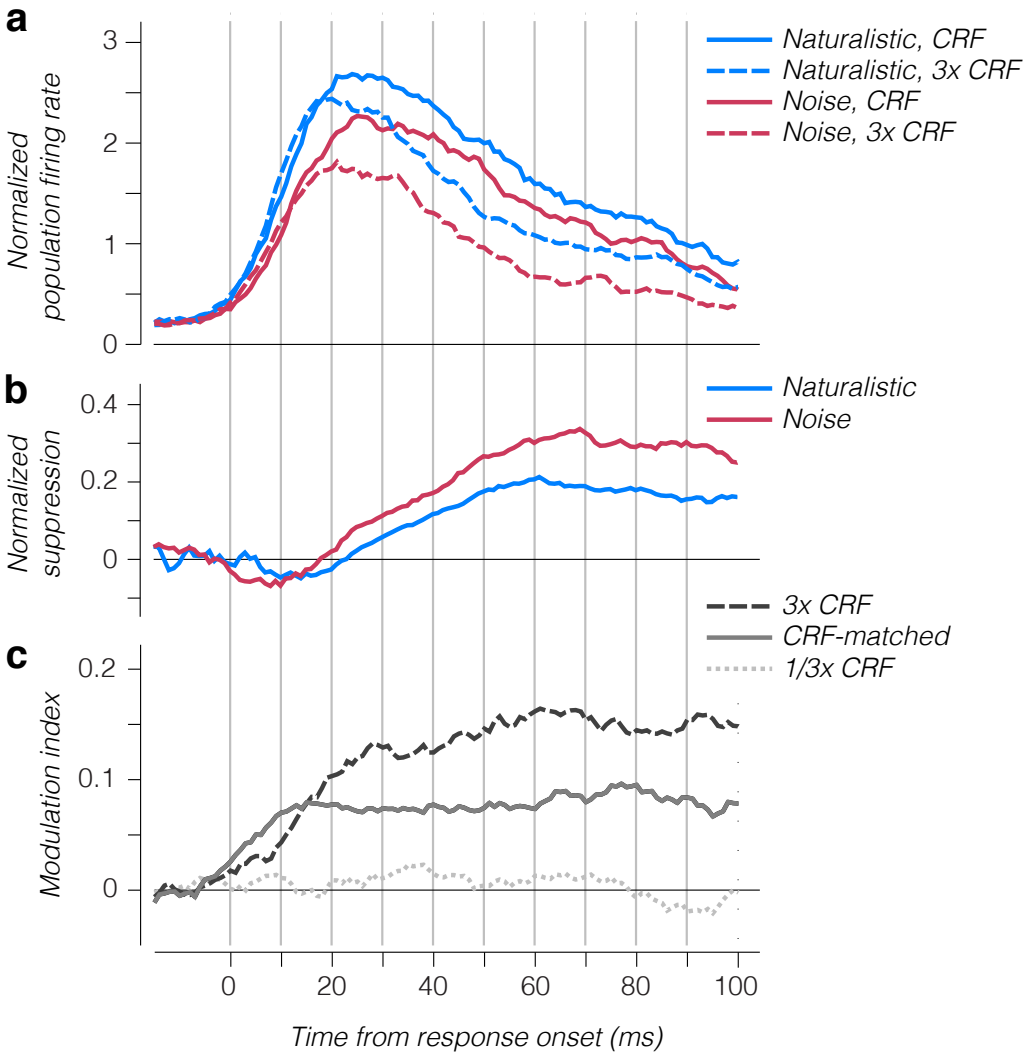

Figure 4.6: Temporal dynamics of suppression and sensitivity. (a) Average normalized population firing of V2 neurons to large and small naturalistic and spectrally matched noise stimuli. The response of each neuron was aligned to its estimated response onset and normalized by its average firing rate during the stimulus period. (b) Normalized suppression as a function of time from response onset for naturalistic and spectrally matched stimuli. (c) Modulation index as a function of response onset for three different sizes.

tions by computing the surround suppression index at each time point (Fig. 4.6b). As has been previously observed [15, 197, 247, 93], we found that suppression from the receptive field surround was delayed, beginning to affect responses about 15–20 ms after response onset and reaching a steady state 50–60 ms post-response (Fig. 4.6b). Surround suppression to spectrally matched noise began a few milliseconds earlier than to naturalistic stimuli and reached a higher value, consistent with our results when aggregating spikes across the entire stimulus window (Fig. 4.5a).

The temporal dynamics of the modulation index had a different but related temporal profile (Fig. 4.6c). For the CRF-matched condition, modulation began to rise right at response onset and reached its steady-state level around 10–15 ms later. When the stimulus was three times the CRF, however, the modulation index continued to rise, surpassing the modulation strength for CRF-matched stimuli around 15–20 ms after response onset. Thus, the time at which surround-enhanced modulation strength became apparent matched the onset of suppressive signals from the surround.

4.4 Discussion

Here, we provide further evidence for the distinctive selectivity of V2 neurons for the statistics of natural textures. The sensitivity of individual V2 neurons for naturalistic visual structure increased as stimuli were extended well beyond the size of the classical receptive field. This observation could partly be explained for some neurons by the recruitment of increased surround suppression for stimuli lacking naturalistic structure. More generally, the delayed onset of surround-enhanced sensitivity and surround suppression in the V2 population was similar, suggesting the possibility that similar mechanisms underlie both phenomena.

The presence of naturalistic statistical dependencies in our stimuli function to either weaken surround suppression, or else drive a facilitative mechanism with similar spatial summation and dynamics. Surround influences on visual neurons have generally been found to be suppressive [21, 43, 115, 34], although facilitative interactions have been observed as well [9, 124, 70]. Can we differentiate between these two possibilities? Firstly, the strength of naturalistic modulation is weakly related or unrelated to the strength of surround suppression over different stimulus ensembles, suggesting different origins. Secondly, neurons with weak or no surround suppression often still exhibited increasing sensitivity to naturalistic structure with size (Fig. 4.2).

Theories for the role of the extraclassical surround have often suggested it acts to decorrelate or increase the sparseness responses to support efficient encoding of visual structure presented to the receptive field [193, 236, 237]. The generally suppressive influence thus acts to weaken responses when the visual structure in the receptive field center and surround are redundant [20]. Surround influences have most often been studied in V1 neurons with simple artificial stimuli in the domain of orientation [115, 43, 203]. When an optimally oriented stimulus is presented to the center, stimulating the surround with the same orientation induces the largest suppression, whereas an orthogonal orientation partially relieves this suppression [203, 35, 247, 93]. The observation that suppression is greatest for matched orientations in the center and surround holds even when the center orientation is nonoptimal, suggesting a general mechanism for suppressing responses to redundant or homogeneous visual elements [203, 35, 38].

This idea has recently been examined and extended using data from V1 neuronal responses to both artificial and natural stimuli [38]. In the proposed frame-

work, surround suppression is gated by a computation that detects visual homogeneity between image content presented to the center and surround of individual neurons. Images containing strong cross-position dependencies, such as some natural images or some of the naturalistic texture families used here, would drive strong suppression, whereas phase-scrambled stimuli with independent structure in the center and surround regions would drive little or no suppression. This conjecture was tested on a population of V1 neurons and surround suppression was indeed found to be weaker for phase-scrambled stimuli compared with an ensemble of natural images [38]. These results match our own small sample of V1 neuronal responses to phase-scrambled and naturalistic texture stimuli (Fig. 4.5b), and suggest a fundamentally different role for the surround in V1 and V2.

Alternatively, our results in V2 may represent a spatially broad facilitative mechanism. Previous suggestions of a facilitatory role for the surround in V1 have emphasized the potential role of long range orientation matched lateral connections that may play a role in contour integration [71, 126, 214, 127]. However, recent work has suggested these signals are significantly delayed and may represent feedback from higher areas induced by perceptual learning [128]. In contrast, the temporal dynamics we observe for the emergence of naturalistic sensitivity are similar to those seen for tuned suppression in V1 thought to originate in lateral connections [115, 15, 197, 93]. Compared with V1 then, our results may indicate a distinctive function for lateral interactions in V2.

The influence of receptive field surrounds has been much less studied in V2 than in V1. When V2 has been studied, surround properties have generally been found to be similar [201], although there may be differences in the spatial layout of suppressive regions in the surround relative to the preferred orientation of the cen-

ter [191]. Interestingly, spatially inhomogeneous surround regions have been linked to the representation of "second-order" features, or texture-defined form, in both V1 and V2 [244, 218, 83]. Receptive field models that incorporate this selectivity can resemble a spatial derivative across spectral features [218, 83]. Depending on its implementation, such a model could potentially yield selectivity to higher-order spatial correlations contained in naturalistic images (See Fig. 2.25). However, previous studies have found little difference in the selectivity to texture-defined form between V1 and V2 [196, 83] (but see [125]).

Naturalistic textures are defined by their repeating elements and broad spatial extent. Limiting the aperture through which one observes a patch of texture can drastically change the qualities of the texture. If texture is made up of visual "stuff," limiting the surrounding elements can allow one to better see the "things" that make it up [6, 5]. Visual neurons are generally thought to view the world through the aperture of their receptive fields. However, our results suggest that V2 neurons are sensitive to the statistical dependencies that determine the appearance of natural visual textures through mechanisms that operate at a scale much larger than individual V2 receptive fields.

Chapter 5

Neuronal signals supporting naturalistic texture discrimination

5.1 Introduction

Different visual behaviors are supported by groups of neurons specialized to encode specific aspects of the visual world. Several brain areas have been linked to specific visual functions through simultaneous characterization of neuronal and behavioral sensitivity [149]. Many studies have found that individual neurons approach or even exceed the perceptual sensitivity of the observer and, moreover, predict behavioral choices even on trials where the visual stimulus is ambiguous [27, 174, 229, 152, 155]. Such results have become a benchmark by which to judge the suitability of a potential neural correlate of behavior.

This framework has been difficult to apply to areas of the brain believed to support midlevel representations of visual form, since we often lack a foundation for exploring their particular visual selectivity. We have recently demonstrated a promising indirect relationship between the response of populations of V2 neurons and the perception of naturalistic textures [58, 260, 261]. Despite being

recorded under anesthesia, V2 responses, but not V1 responses, predicted human psychophysical performance on a naturalistic texture discrimination task. Here, we test the strength of this link between V2 neurons and the perception of naturalistic image structure by measuring neuronal and behavioral sensitivity simultaneously in the same observer.

We found that V2 neurons were much more sensitive to naturalistic image structure than V1 neurons, confirming our previous results [61]. However, average sensitivity in V1 and V2 was far from behavior. Despite this, we found a significant relationship between neuronal responses and choice in both V1 and V2. The temporal profile of this choice-related activity was markedly different in V1 and V2. In V2, neurons were both selective for the stimulus and predictive of upcoming choice soon after stimulus onset. In contrast, V1 neurons gradually built-up stimulus selectivity and choice-related activity until they peaked just before the choice. Our results differentiate the functional role of V1 and V2 neurons in the perception of naturalistic visual structure, and also demonstrate the emergence of neuronal signals supporting naturalistic texture discrimination across both time and different brain regions.

5.2 Methods

5.2.1 Physiology

Two male macaque monkeys (1 *M. mulatta*, 1 *M. nemestrina*) were trained to perform a naturalistic texture discrimination task and an orientation discrimination task. Neuronal recordings were only made from one animal during performance of the texture discrimination task, so we report results from only this subject.

Experimental procedures conformed to the National Institute of Health Guide for the Care and Use of Laboratory Animals and were approved by the New York University Animal Welfare Committee. Under general anesthesia, the animal was implanted with a titanium head post and recording chamber [2, 3]. Extracellular recordings were made with dura-penetrating glass-coated tungsten microelectrodes (Alpha Omega), advanced mechanically into the brain. We distinguished V1 from V2 on the basis of depth from the cortical surface and changes in the receptive field location of the recorded units. We made recordings from every single unit with a spike waveform that rose sufficiently above noise to be isolated. We first presented suitably vignetted sinusoidal grating stimuli to map each isolated unit's receptive field in a fixation task. Across our population, receptive fields were centered at eccentricities ranging from 2° to 6°. Thereafter, we ran an initial characterization of each neurons sensitivity to naturalistic image structure. We presented a random sequence of different samples of naturalistic and spectrally matched noise stimuli from five different families. Each image was presented for 100 ms with 100 ms of intervening mean luminance. We computed the discriminability between the average firing rate across samples of naturalistic stimuli and spectrally matched noise for each of the five families. We then selected the the family for which naturalistic and spectrally matched noise were most discriminable for use in the texture discrimination task.

5.2.2 Behavioral task

We generated our naturalistic texture and spectrally matched noise stimuli as described previously in section 4.2.1. Stimuli of intermediate naturalness were synthesized as described in section 2.2.1. Subjects were seated in a semi-dark room

in front of a gamma-corrected CRT monitor (iiyama HM204DTA) with their heads stabilized. Eye position was recorded with a high-speed, high-precision eye tracking system (EyeLink 1000). We presented visual stimuli at a viewing distance of 57 cm, a spatial resolution of 1,280 × 960 pixels, and a refresh rate of 120 Hz. Stimuli were presented using Expo software (`http://corevision.cns.nyu.edu/expo/`) on an Apple Macintosh computer.

Each trial in the discrimination task began when subjects fixated a small white point at the center of the screen (0.2° diameter). After 250 ms, two choice targets appeared, one on each side of the fixation point (on the horizontal meridian, at 3.5° eccentricity). For the texture discrimination task the choice targets were samples of naturalistic and spectrally matched noise presented in a 2° aperture that did not change over the course of a session. In the orientation discrimination task, the choice targets were white lines (0.3° wide, 2.0° long) rotated –22.5° and 22.5° away from the discrimination boundary. After a 500 ms delay, the target appeared. In the texture discrimination task, the target was a sample of texture of intermediate naturalness presented within a 4° aperture. In the orientation task, the target was a drifting grating. The stimulus remained on for 500 ms. In the texture discrimination task subjects judged whether the target was more similar to naturalistic or spectrally matched noise. In the orientation task, subjects judged the orientation of the stimulus relative to the discrimination boundary. When the stimulus disappeared, the fixation point turned black and subjects reported their decision with a saccadic eye movement to one of the choice targets. If the monkeys made a saccade to the correct choice target, they received a liquid reward.

We varied naturalness over a range from fully naturalistic (naturalness = 1) to fully phase-randomized (naturalness = 0). We varied orientation over a 30° range

centered on the discrimination boundary. Stimuli were presented in random order. On each trial, naturalistic texture stimuli were drawn from 300 different possible samples within each naturalness condition. New samples were generated for each texture family once a week to prevent the animal from memorizing individual samples. However, on a subset of tasks in the ambiguous condition where naturalness = 0.5, we overrepresented a single sample to measure behavioral and neuronal covariability without the added variance from the presentation of different samples. Stimuli that were matched to the discrimination boundary were rewarded randomly. Trials in which the subject did not maintain fixation within 0.6° of the fixation point were aborted. Data are reported from every experiment for which at least 100 trials were completed.

5.2.3 Analysis of behavioral response

We measured the observer's behavioral capability to discriminate the naturalness of texture stimuli by fitting the relationship between stimulus naturalness and probability of naturalistic choice with a psychometric function consisting of a lapse-rate and a cumulative Gaussian function. We optimized model parameters by maximizing the likelihood over the observed data, assuming responses arise from a Bernoulli process. We defined the decision criterion and estimation noise as the mean and standard deviation of the cumulative Gaussian. In the vast majority of experiments, lapses were rare, indicating that observers' responses were almost always informed by the stimulus. Experiments in which the lapse rate exceeded 15% were excluded from the population analysis.

5.2.4 Analysis of neuronal response

We computed each neuron's stimulus response by counting spikes in a 500 ms window following response onset. For each cell, we chose a latency by maximizing the stimulus-associated response variance [212]. We measured strength of neuronal tuning in the orientation discrimination task by computing the correlation between (relative) stimulus orientation and response mean. For orientation-selective neurons with a suitably chosen discrimination boundary, the absolute value of this correlation approximates one. Small values occur when orientation selectivity is very weak, or when the discrimination boundary is misplaced. Data are reported from every unit for which the absolute value exceeded 0.5. The strength of neuronal tuning to texture naturalness was much weaker. Here we measured the sign of tuning by conducting a logistic regression analysis between the spike counts of individual neurons and the animals choices across different conditions. The slope of this function reflects the strength and sign of tuning. We performed analyses on all neurons for which we collected sufficient data and the animal performed adequately, regardless of the strength of tuning.

We measured neuronal capability to discriminate stimuli by fitting the relationship between stimulus naturalness (or orientation) and probability of naturalistic (or "clockwise") choice for an ideal observer with a cumulative Gaussian function. The ideal observer's choices were obtained by applying a deterministic decision criterion to the responses of each neuron. To minimize bias in the ideal observer's choices, the criterion was set to the median response to the zero signal stimulus. We defined neuronal estimation noise as the standard deviation of the cumulative Gaussian. For each cell, we computed choice probability (CP) for the zero-signal

stimulus and the neighboring stimulus conditions. As described previously [26], CP is calculated by performing an ROC analysis on the choice-conditioned neuronal responses to repeated presentations of a single stimulus. We classified a CP estimate as statistically significant if it fell outside of the central 95% of the expected null distribution, computed from 1000 randomly permuted data sets [26].

We computed average peristimulus time histograms (PSTHs) by counting spikes in 1 ms windows for individual neurons, normalizing to the maximum observed firing rate, and averaging across neurons. Population PSTHs were then smoothed using a causal exponential filter with a time constant of 25 ms. To compute the dynamics of CP we calculated CP for individual neurons within 100 ms time windows each shifted by 10 ms before averaging together CP values across the population.

5.3 Naturalistic texture discrimination

We trained macaque monkeys to discriminate the "naturalness" of a peripherally presented patch of texture (Fig. 5.1). On each trial, after the animal attained stable fixation, we presented two choice targets to the left and right of fixation in the upper visual field. One choice target was a naturalistic texture matched to an original natural photograph for the joint statistics of the outputs of V1-like filters tuned to different orientations, spatial frequencies, and positions. [172]. The other choice target was phase-randomized noise, spectrally matched to the same original photograph, and lacking the higher-order statistics of the naturalistic texture. Following the choice targets, we presented a stimulus centered on the receptive field of a simultaneously recorded single unit in V1 or V2. We generated the target stimulus by titrating the strength of higher-order statistics to create intermediate levels of naturalness. When no higher-order statistics were included

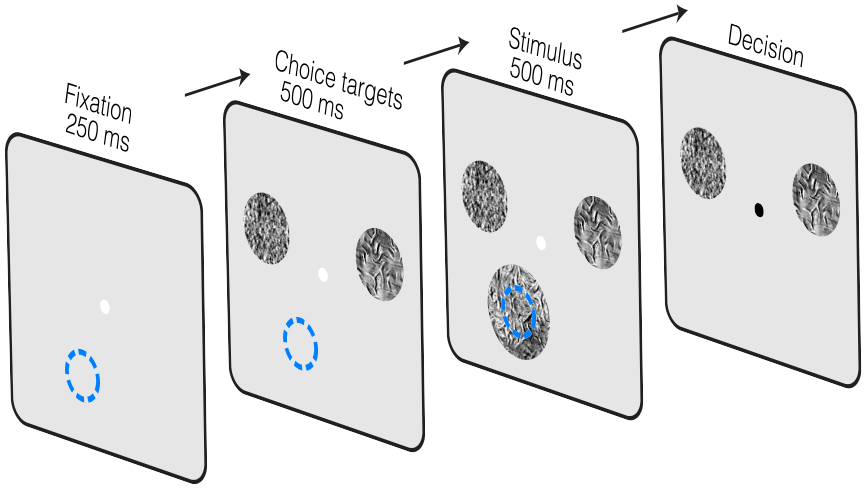

Figure 5.1: Naturalistic texture discrimination task. The animal attained fixation on a central point. After 250 ms two choice targets appeared, one naturalistic texture and one spectrally matched noise. 500 ms later a target stimulus appeared centered on the receptive field of a simultaneously recorded V1 or V2 neuron. 500 ms later, the animal was prompted to indicate its decision with a saccade to one of the choice targets. Blue circle represents a receptive field and was not part of the visual display.

in the synthesis (naturalness = 0) the stimulus was spectrally matched noise, and when the higher-order statistics were fully imposed (naturalness = 1) the stimulus resembled naturalistic texture. The monkey indicated whether the target was more naturalistic or noise-like by making a saccade to one of the choice targets after stimulus offset. Target stimuli above 0.5 naturalness were rewarded for a saccade to the naturalistic choice target. Stimuli at exactly 0.5 were rewarded randomly.

The monkey performed naturalness discrimination for five texture families generated from different original photographs. The five naturalistic texture families corresponded to those producing the largest differential response in a population of V2 neurons compared with spectrally matched noise (top five in Fig. 2.5). At the beginning of each session, we isolated a single unit in V1 or V2, mapped the center

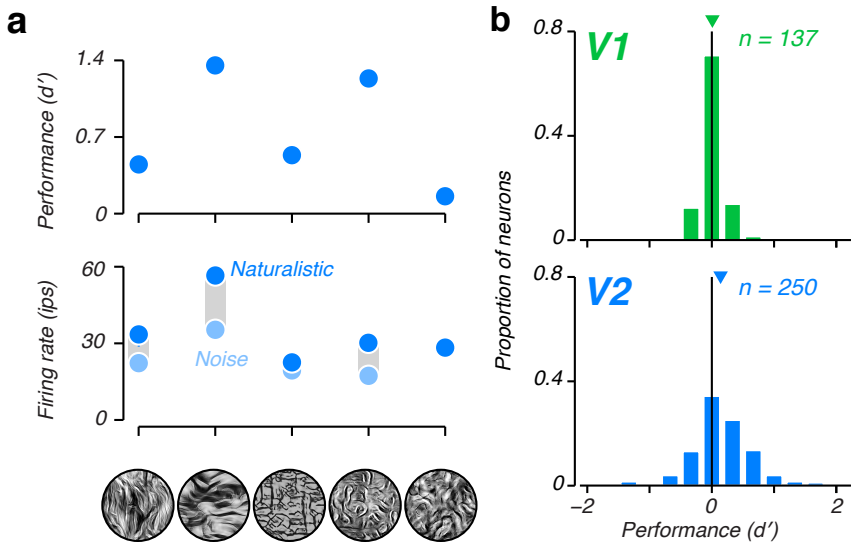

Figure 5.2: Tuning to naturalness in V1 and V2. (a). The firing rate of an example V2 neuron to naturalistic and spectrally matched noise stimuli drawn from five texture families (bottom). Corresponding discriminability (d′) values for each texture family obtained form an ideal observer analysis applied to the spike counts (top). (b) Distribution of average signed d′ of single neurons in V1 and V2. Negative d′ values indicated the neuron preferred noise stimuli over naturalistic stimuli. Mean d′ was -0.013 in V1 and 0.13 in V2.

of its receptive field and briefly determined its tuning to naturalness (Fig. 5.2a, bottom). We presented multiple samples of naturalistic and spectrally matched noise images from the five families, each presented for 100 ms and separated by 100 ms of gray screen. We calculated the discriminability between the distribution of spike counts to the two categories of stimuli (Fig. 5.2a, top). We selected the texture family that evoked the most discriminable responses between naturalistic and noise for the subsequent behavioral task, regardless of whether the neuron preferred naturalistic or noise stimuli. By choosing the most neurally discriminable of five families, we attempted to optimize the discrimination task the animal performed to rely on the neuron under study. However, our optimization is admittedly limited by the small number of texture categories in comparison with the probable

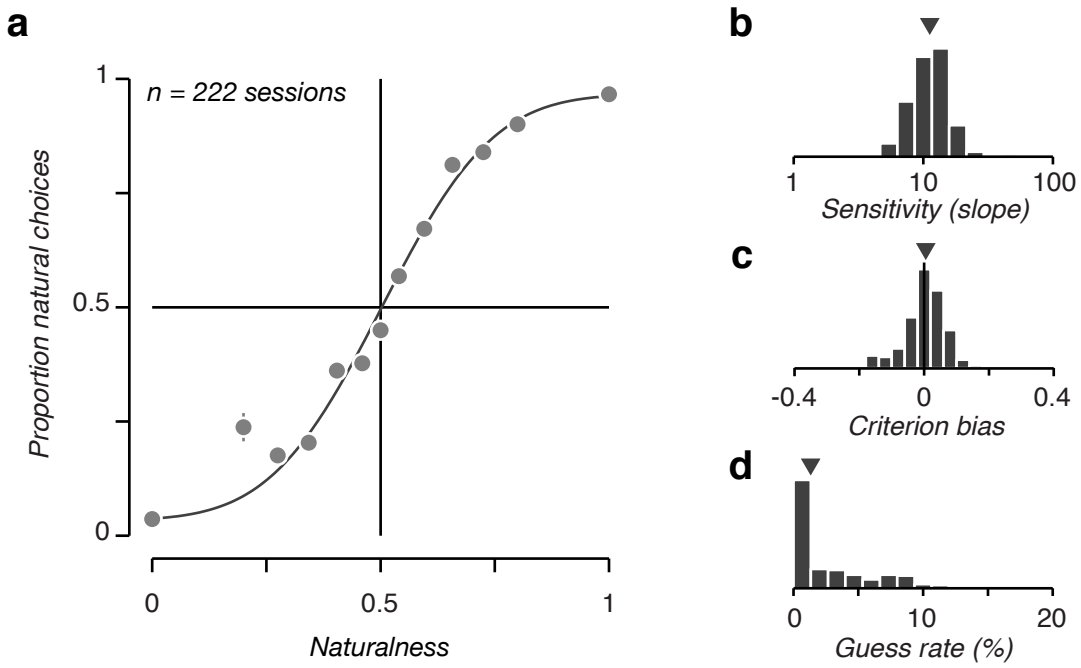

Figure 5.3: Behavioral performance over many sessions. (a) Average psychometric function across all sessions. The points represent measured behavior. The line represents a fit of the signal detection theory model of behavior. (b) Distribution of sensitivity over all sessions. (c) Distribution of bias parameter over all sessions. (d) Distribution of guess rate over all sessions.

complexity of selectivity for naturalistic image structure.

Examining the average signed discriminability over all five texture families for the entire population of recorded neurons, we found a very similar pattern to that observed in anesthetized macaque cortex (Fig. 5.2b, compare with Fig, 2.7). Specifically, there was no overall preference in the V1 population for naturalistic or spectrally matched noise stimuli, whereas V2 neurons were driven to higher firing rates by naturalistic stimuli.

The subject performed the task well. When the target had a naturalness value of 0 or 1 they were nearly perfect, and performance progressively declined as naturalness approached the bound of 0.5 (Fig. 5.3a). Naturalness of 0.5 is the midpoint

between naturalistic and noise stimuli for the model parameters, and we enforced this discrimination bound through the structure of the rewards. However, o.5 does not necessarily correspond to the perceptual midpoint and so the subject could potentially develop biases in its judgments. To distinguish perceptual sensitivity from bias, we fit the animal's behavior with a model in which choices arise from applying a deterministic decision criterion to a noisy estimate of naturalness. We transformed the estimation noise into a measure of sensitivity, which corresponds to the slope of the psychometric function at the discrimination bound. The subject had high sensitivity (Fig. 5.3b) and little bias (Fig. 5.3c) despite having to learn the 0.5 naturalness bound for each family.

5.4 Neuronal and perceptual sensitivity

Although the subject learned to perform the task well, and was quite consistent across sessions, the sensitivity of single neurons to naturalness varied widely. To estimate neuronal sensitivity we performed an ideal observer analysis on the distribution of spike counts to different levels of naturalness. We directly compared the resulting neurometric functions to the simultaneously recorded psychometric function (Fig. 5.4). In some cases, the sensitivity of a single neuron approached the sensitivity of the entire organism (Fig. 5.4a). Neurons with sensitivity approaching that of behavior were much more common in V2 than V1. However, we more typically found that the sensitivity of the monkey far exceeded that of single neurons in both V1 (Fig. 5.4c) and V2 (Fig. 5.4b).

We found that over the whole population of recorded units, sensitivity in V2 was significantly greater than in V1 (Fig. 5.5; $P < 0.0001$, Wilcoxon rank sum test). On average, V1 neurons were 18 times less sensitive than the animal's behavior

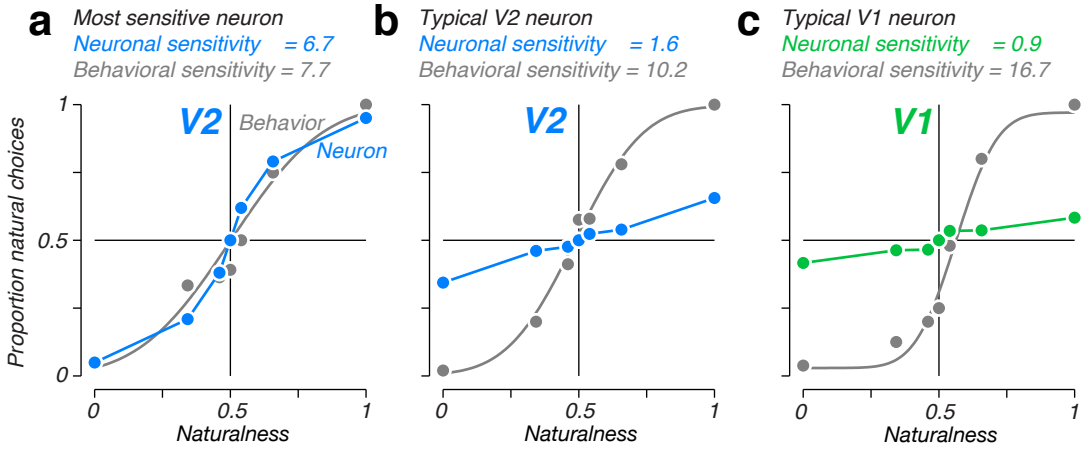

Figure 5.4: Example sessions. (a) Neurometric and psychometric functions obtained during recording a highly sensitive V2 neuron. Gray symbols represent measured behavior, and the gray line represents the model fit to these data. Colored symbols represent ideal observer analysis applied to the responses of a single V2 neuron. Neuronal sensitivity was obtained by fitting the signal detection theory model to this neurometric function and, extracting sensitivity, and dividing by the square root of 2 to correct for the added sensitivity of the "anti-neuron" (see methods). (b) Another example V2 cell, much less sensitive than the animal. (c) An example V1 cell even less sensitive than behavior.

while V2 neurons were 11 times less sensitive than behavior. In V1, neurons were roughly equally split between those that preferred naturalistc and those that preferred spectrally matched noise stimuli (Fig. 5.5), and there was no difference in the average sensitivity of these two populations ($P = 0.58$). In contrast, in V2 there were twice as many units preferring naturalistic over noise stimuli and this population was over twice as sensitive (Fig. 5.5; $P < 0.0001$). The population of V2 neurons that preferred spectrally matched noise was on average 18 times less sensitive than behavior and was not significantly different from the V1 population ($P = 0.40$). The population of V2 neurons that preferred naturalistic textures was only 8 times less sensitive than behavior (Fig. 5.5).

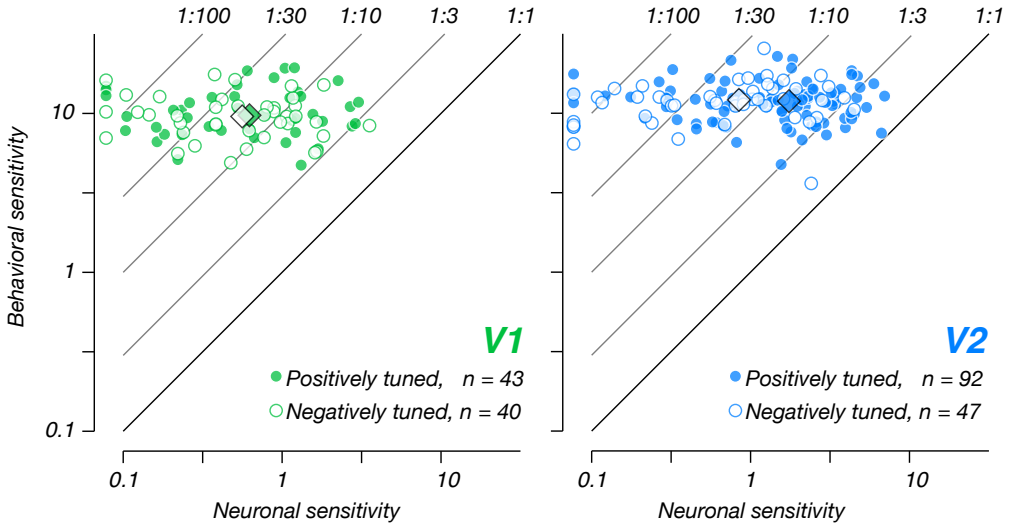

Figure 5.5: Perceptual and neuronal sensitivity to naturalness. Estimated sensitivity of single neurons plotted against the simultaneously obtained behavioral sensitivity for V1 and V2 neurons. Neurons are grouped by those that preferred naturalistic over spectrally matched noise stimuli (Positively tuned) and those that had opposite preferences. Black outlined diamonds represent the geometric means of each population.

5.5 Comparison with orientation sensitivity

To put these sensitivity numbers in perspective, we compared the naturalistic texture sensitivity of V1 and V2 neurons to that of a more well known form of visual cortical selectivity. Prior to training on the naturalistic texture discrimination task, we previously trained the same monkey to perform a fine orientation discrimination task (Fig. 5.6). The orientation task followed the same temporal structure as the texture discrimination task. Instead of texture patches, in this case the choice targets were two oriented bars, rotated 45°apart. The discrimination boundary was implicitly indicated to the animal by the orientation midway between the orientation of the two choice targets. In this task, the discrimination boundary was fully optimized for the neuron under study. We measured the orientation tuning

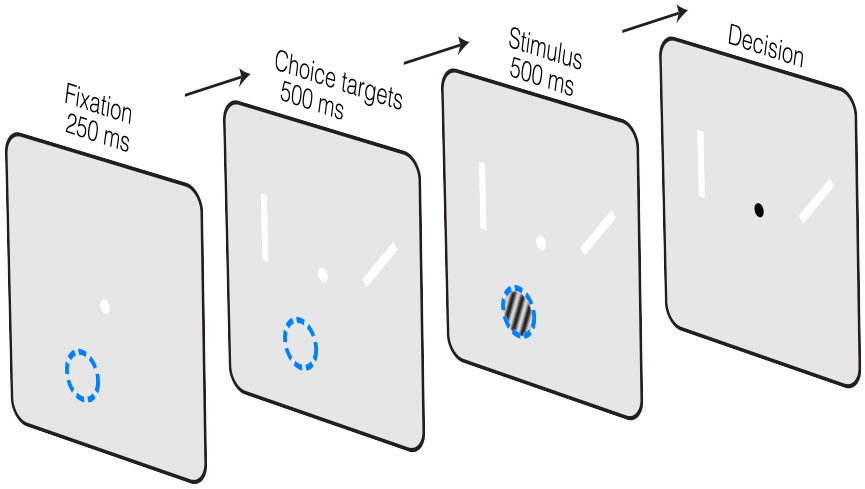

Figure 5.6: Orientation discrimination task. The animal attained fixation on a central point. After 250 ms two choice targets appeared in the form of bars oriented 45° from each other. The orientation midway between them was the orientation discrimination bound. 500 ms later a target stimulus appeared centered on and sized optimally for the receptive field of a simultaneously recorded V1 or V2 neuron. 500 ms later, the animal was prompted to indicate its decision with a saccade to one of the choice targets. Blue circle represents a receptive field and was not part of the visual display.

of each neuron prior to the behavioral experiment, and placed the orientation boundary on the steepest part of the tuning curve.

Under these conditions, neuronal sensitivity was much closer to that of behavior. To quantify this, we performed the same ideal observer analysis on these data that we applied to the naturalness task. On average, V1 neurons were only 1.75 times less sensitive than the behavior of the entire organism, and V2 was only 1.6 times worse (Fig. 5.7). We found no discernable difference in the sensitivity of V1 and V2 to orientation ($P = 0.81$). Thus, V2 and especially V1 appear much more selective for orientation than for naturalistic image structure, but the difference was likely compounded by the inability to optimize the stimulus fully for individual neurons in the texture discrimination task.

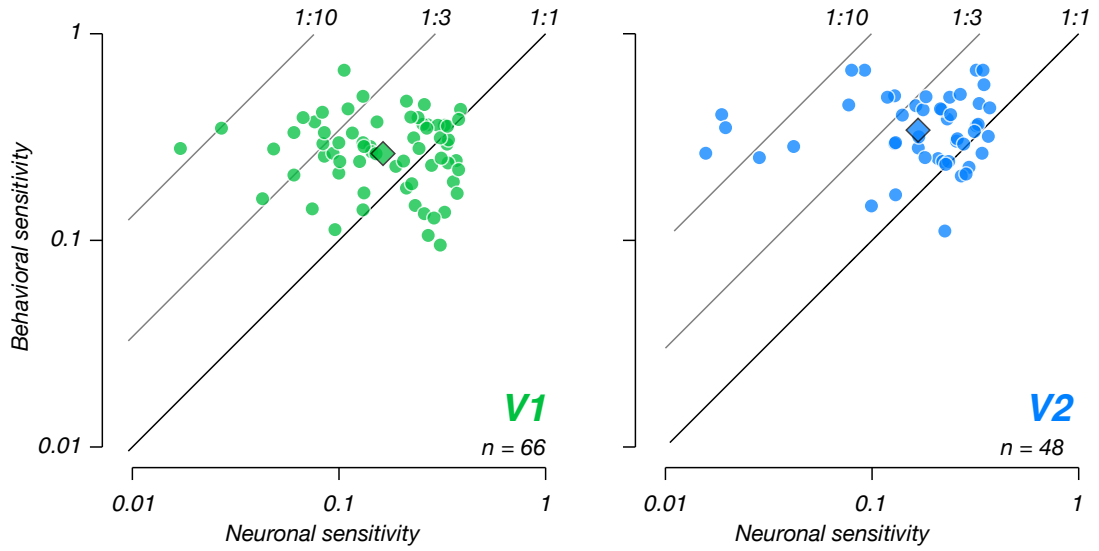

Figure 5.7: Perceptual and neuronal sensitivity to orientation. Estimated sensitivity of single neurons plotted against the simultaneously obtained behavioral sensitivity for V1 and V2 neurons. Black outlined diamonds represent the geometric means.

5.6 Choice-related signals in V1 and V2

Given the large disparity between neuronal and perceptual sensitivity in the texture discrimination task, we wondered whether we could find further direct evidence for the causal involvement of these neurons in perceptual decisions. In particular, we wondered whether the increased sensitivity of V2 neurons would manifest in a greater tendency to predict perceptual decisions on a trial-by-trial basis. To examine this, we computed "choice probability" for the responses of each neuron to the ambiguous 0.5 naturalness condition. This quantity measures the probability that a neuron fires more spikes preceding a behavioral decision associated with its preferred stimulus. Although choice probability values varied widely across the population of individual neurons, the population was shifted significantly above the chance value of 0.5 in both V1 (mean = 0.54; $P < 0.005$, t test) and V2 (mean

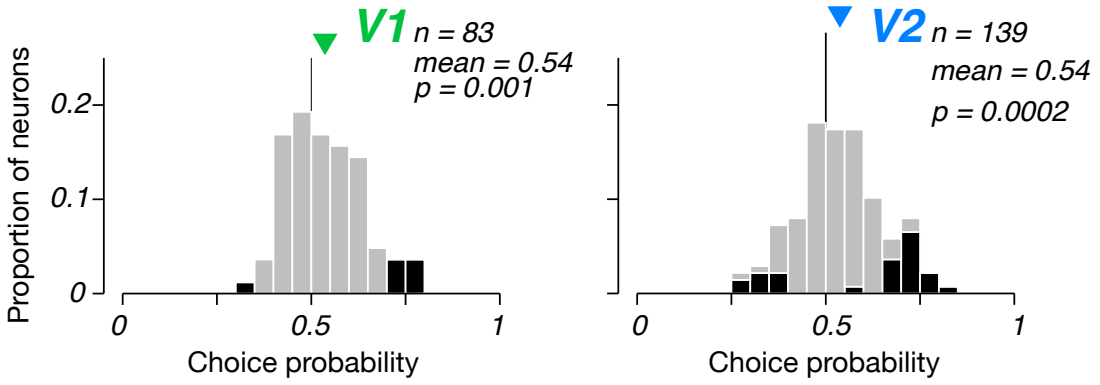

Figure 5.8: Choice probability for naturalistic texture discrimination. Distribution of choice probabilities computed using responses to the ambiguous stimulus condition for V1 (left) and V2 (right) neurons. Black bars indicate neurons that were significantly different from 0.5 by the permutation test introduced by [26]

= 0.54; P $<$ 0.0005; Fig. 5.8).

These values of choice probability are weak but not substantially different from values often reported in the literature [26, 152, 155]. However, our experimental paradigm differs from many previous experiments for which significant choice probabilities have been observed. Specifically, we presented a single static image for 500 ms during a trial in contrast to most studies that have presented dynamic noise stimuli. Our previous results indicate that differences in statistically matched

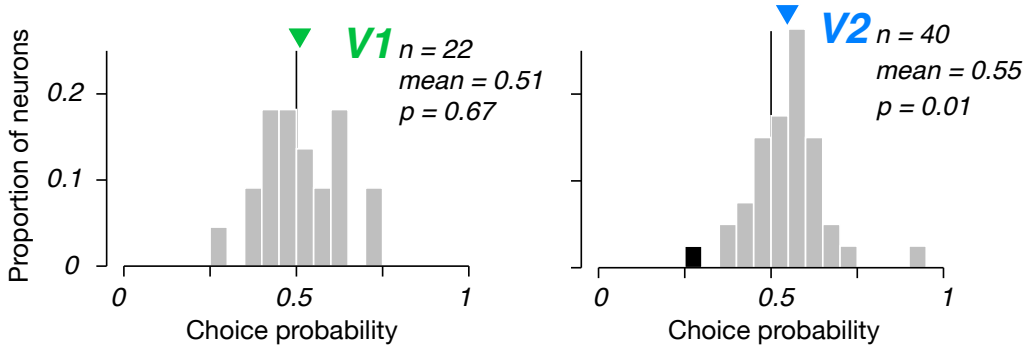

Figure 5.9: Distribution of choice probabilities computed from responses to identical samples at 0.5 naturalness for V1 (left) and V2 (right) neurons.

texture samples account for a substantial amount of neuronal variability in V1 and V2 (Fig. 3.2, [260]). Variability from different samples could thus influence behavioral reports as well as neuronal activity and lead to spurious correlations. To examine whether this possibility contributed to our observations, we displayed a single sample of the 0.5 naturalness condition multiple times during a subset of experiments. We selected experimental sessions where the animal made differing behavioral responses to multiple presentations of the same sample and analyzed choice probability. We found that although choice probability was no longer significantly different than chance in V1 (mean = 0.51; P = 0.67), choice probability was largely unchanged in V2 (mean = 0.55; P < 0.05; Fig. 5.9). However, there was no significant different between choice probability in V1 and V2 (P = 0.21).

We were struck that sensitivity differences were so strong between V1 and V2 but choice probability was not substantially different. Previous studies have consistently found that the most sensitive neurons tend to have the strongest association

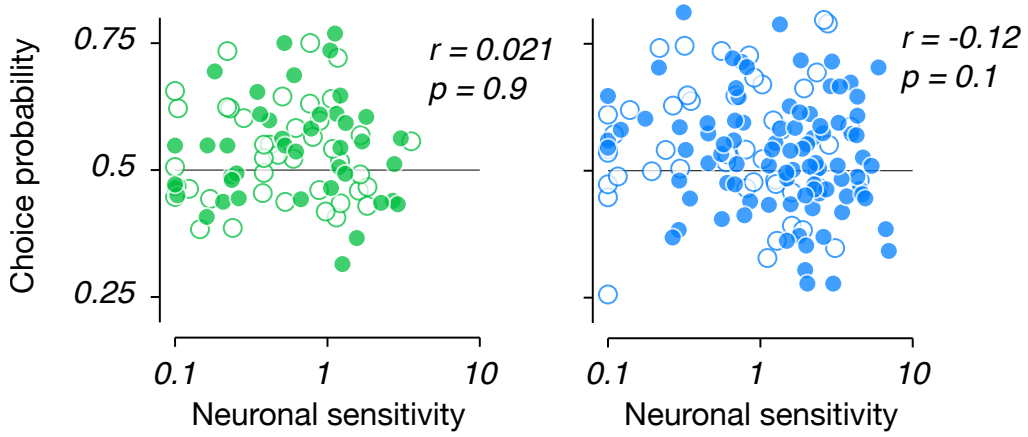

Figure 5.10: Neuronal sensitivity plotted against choice probability for V1 (left) and V2 (right) neurons. Filled symbols represent neurons that preferred naturalistic stimuli, and open symbols represent neurons that preferred spectrally matched noise stimuli. There was no significant correlation in either area.

with choice. However, we found no evidence for a relationship between neuronal sensitivity and choice probability in either V1 or V2 (Fig. 5.10), consistent with the lack of a relationship across areas.

5.7 Dynamics of neuronal responses in V1 and V2

We wondered whether a more detailed analysis of the time course of responses in V1 and V2 could shed light on the apparent inconsistency between sensitivity and choice-related activity across V1 and V2. We first examined the average population firing rate across our population of V1 and V2 neurons (Fig. 5.11). In both areas, there was a prominent initial transient in firing rate following stimulus onset. The transient was stereotyped in its amplitude across different naturalness conditions, indicating that this initial portion of the response contained little stimulus selectivity. To quantify the dynamics of population selectivity, we computed the variance across the average firing rate to each condition, yielding a measure of the

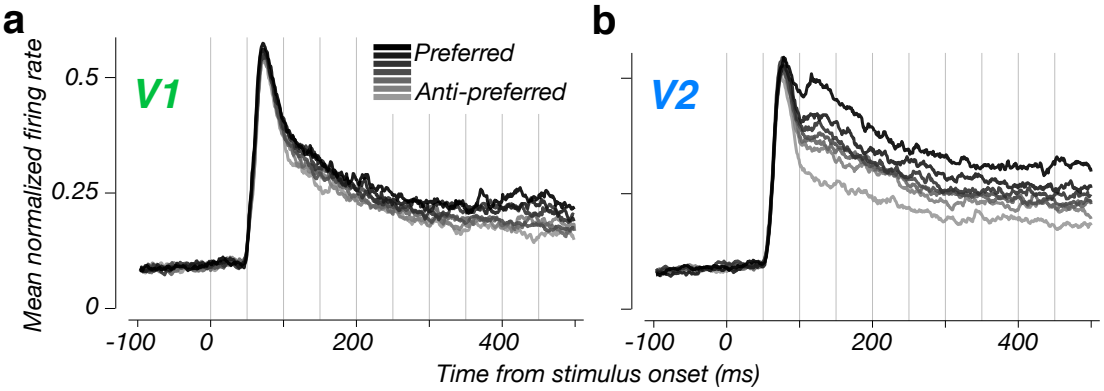

Figure 5.11: Normalized population firing rate to different levels of naturalness in V1 (left) and V2 (right). The firing rate of each neuron was normalized by its mean firing rate before averaging. The stimulus conditions for neurons that preferred spectrally matched noise were flipped before averaging so that the black lines represent the firing rate to the preferred stimulus (either naturalistic stimuli or spectrally matched noise). The stimulus was on for 500 ms.

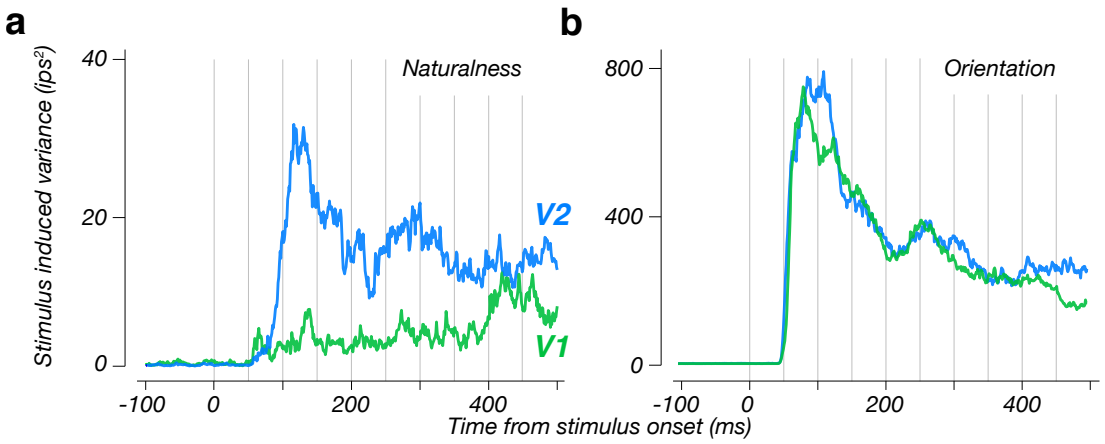

Figure 5.12: Dynamics of selectivity for naturalness and orientation. We plotted the stimulus-induced variance across the population of V1 and V2 neurons for naturalistic stimuli (right) and orientation stimuli (left). We computed stimulus-induced variance by computing the variance across the average response to each condition (Fig. 5.11. Note different ordinate scales for naturalistic stimuli and orientation stimuli.

stimulus induced variance.

The dynamics of this selectivity measure differed markedly across V1 and V2 (Fig. 5.12a). The stimulus-induced variance was very weak in V1 and developed only gradually, peaking over 400 ms after stimulus onset. In contrast, variance in V2 peaked early (between 100 and 150 ms) and subsequently maintained a steady level throughout the stimulus period. This difference between V1 and V2 was specific to naturalistic texture stimuli. When we examined the time course of orientation-induced variance there was no prominent difference in either overall selectivity or the dynamics of selectivity between V1 and V2 (Fig. 5.12b). However, even in V2, which showed much stronger selectivity for naturalness than V1, the orientation-induced variance was more than an order of magnitude larger than that induced by varying naturalness.

Interestingly, although selectivity for naturalness peaked much earlier in V2 than in V1, the peak was quite delayed with respect to response onset. This

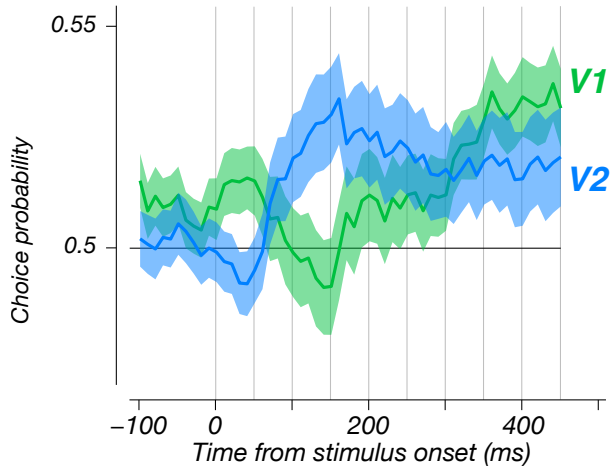

Figure 5.13: Dynamics of choice probability. Average choice probability across the population of V1 and V2 neurons was computed in 100 ms time bins shifted by intervals of 10 ms. Error bars represent ±s.e.m. across the population.

reflected the presence of the non-selective transient response in the population firing rate (Fig. 5.11b). We found that this was not a feature of selectivity for orientation. Selectivity for orientation in both V1 and V2 appeared to peak roughly 50 ms before it did for naturalness in V2. Thus, while orientation information appears immediately in the firing rate of individual neurons, the first spikes in V2 don't appear to carry information about the naturalness of the stimulus.

Given the markedly different dynamics of selectivity between V1 and V2, we wondered whether choice-related activity would also differe between the two areas. We computed choice probability within 100 ms windows shifted by 10 ms for each individual neuron and then averaged the result. We found that choice-related activity evolved over a different time course in V1 and V2 (Fig. 5.13). In V1, choice probability was roughly 0.5 for the first 150 ms following response onset. Choice probability began to rise later during the stimulus period and peaked just before stimulus offset — a temporal profile that resembled the lack of early selectivity in V1. In V2, choice probability grew to be larger than 0.5 soon after response onset, peaking at about 150 ms post-stimulus and remaining high thereafter. Thus,

the distinct dynamics of choice-related activity in V1 and V2 appeared to mirror differences in the dynamics of selectivity in the two areas.

5.8 Discussion

We have uncovered multiple functional differences between V1 and V2 neurons active during the performance of a naturalistic texture discrimination task. We found much greater sensitivity for the higher order structure present in natural texture images, replicating our previous findings differentiating V1 and V2 neurons recorded from the anesthetized macaque cortex [61, 260]. We further show that despite the increased sensitivity from V1 to V2, V2 neurons are still much less sensitive than behavior on average. This finding leaves open the possibility that although sensitivity to naturalistic image structure first emerges in V2, it is further consolidated downstream, potentially in V4 [12, 184, 162]. Recent work has demonstrated selectivity in V4 neurons for the higher-order statistics present in our naturalistic stimuli [162]. A direct comparison between V2 and V4 suggests that selectivity for higher-order statistics may be increased relative to the representation of spectral statistics in V4 [161]. fMRI results in humans also suggest that areas downstream of V2, especially V3, contain similar or increased sensitivity to naturalistic image structure (Fig. 2.12).

Although the finding of increased naturalistic selectivity of V2 over V1 neurons is consistent with our previous work, there are some notable differences. First, the strength of sensitivity to naturalistic statistics appears to be slightly lower than in our anesthetized experiments (compare Fig. 5.5b and Fig. 2.7). Similarly, the number of V2 neurons that showed a preference for spectrally matched noise over naturalistic stimuli was much higher. Although this inconsistency could conceiv-

ably be a consequence of recording neuronal activity during the awake versus the anesthetized state, it is more likely a simple sampling issue. The current data are from just one monkey, whereas our previous experiment contained data from thirteen. There is some evidence that neural sensitivity to naturalistic image structure develops over the course of months in infancy [243], and can be disturbed by abnormal visual development [262]. It is thus possible there are simply differences in the strength of sensitivity in V2 across individuals. Another possibility is that the constraints on awake physiological recording caused us to oversample one of the anatomically-defined compartments in V2, and it is currently unknown whether sensitivity to naturalistic image structure differs across these compartments.

We also observed a markedly different temporal emergence of sensitivity to naturalistic statistics compared with our previous work (compare Fig. 2.4c and Fig. 5.12a). In this case, a large portion of this difference likely reflects the effect of anesthesia on subcortical and cortical responses [8, 107]. In our anesthetized recordings there was no initial transient to stimulus onset in the population activity of V1 or V2, but a gradual rise in responses. This could be due both to slower dynamics within individual neurons [107], or more variability in the latency to response across neurons. In the anesthetized state, sensitivity to naturalistic statistics was present at response onset and gradually develop over the course of tens of milliseconds. Here, there was a significant delay in the onset of naturalistic sensitivity (Fig. 5.12a). This observation, along with our previous finding that the onset of surround enhanced sensitivity is delayed (Fig. 4.6), may suggest that intracortical processing within V2 (or potentially feedback from higher areas) plays a significant role in establishing selectivity for naturalistic image structure. The dynamics we report here also differ from the previously reported fast onset of

"shape selectivity" recorded in awake V2 [90]. However, this form of selectivity does not diverge markedly from that in V1, and so may be inherited [91]. Along these lines, the dynamics of naturalistic sensitivity differ strikingly from the emergence of orientation selectivity (Fig. 5.12b), a property known to be influenced largely by linear filtering of feedforward inputs [173, 75]. In contrast, a functional account of naturalistic sensitivity requires two stages of processing in V2 (Fig. 2.25a), and the slow emergence of selectivity may reflect the dynamics of this computation.

Although we found much higher sensitivity for V2 neurons compared with V1, average sensitivity in V2 neurons was still very far from behavioral performance. Many studies have found a tight correspondence between neuronal and behavioral thresholds [27, 174, 229, 152, 155], and suggest a clear role for the area under study in supporting a particular perceptual behavior. However, each of these studies, as well as our own orientation discrimination experiment, involved tailoring the target stimulus and discrimination bound to the tuning preference of each individual neuron. This procedure gives the single neuron the best chance to match behavioral performance. We currently do not have a detailed understanding of the space of naturalistic statistics and cannot perform tuning experiments along meaningful axes of selectivity (although see [162] for an example of adaptively searching the parameter space). Instead, we are limited to characterizing the differential response for a predefined set of texture families (for which the animal has been previously trained) and performing subsequent behavioral experiments using the most discriminable texture family for the recorded neuron. Compared with the hundreds of parameters in the texture model used to generate our naturalistic stimuli [172], five texture families is not many for truly optimizing the stimulus. Had we been able to use more texture families in characterizing neuronal selec-

tivity, it's very probable we would have found a closer correspondence between neuronal and behavioral sensitivity.

Moreover, the naturalistic sensitivity of V2 neurons compared with behavior is comparable to previous studies that have not optimized stimuli to the tuning preferences of recorded neurons [175, 79, 80, 198]. Average neuronal sensitivity in these studies can range from 3 to 17 times less sensitive than behavior. The pattern observed in most of these studies is a very wide distribution over neuronal sensitivity, with the most sensitive neurons approaching or overtaking behavioral sensitivity. Broadly, this is the pattern of sensitivities we see in V2, although V1 had very few neurons that approached behavioral sensitivity. Thus, Our observation of inconsistent neuronal and behavioral sensitivity does not necessarily indicate V2 is an inappropriate neural substrate for supporting naturalistic texture discrimination.

Many studies, both those that optimize stimuli for single neurons and those that don't, have found significant choice-related activity. However, the interpretation of this observation has grown murkier over the last decade. It was initially suggested that the inherent noise in sensory neurons causally affected the decision process downstream, manifesting as correlations between single sensory neurons and behavior [26, 194, 170]. Recent results have instead posited a feedback origin of such correlations, whereby the outcome of a perceptual decision is relayed to early sensory neurons [154, 151]. The temporal emergence of choice probability has been important in disentangling contributions from feedforward versus feedback sources. Here we report two very different temporal profiles of choice-related activity in V1 and V2. The dynamics of choice-related signals in V2 are similar to many previous reports in the literature, which show an early onset of significant

choice probability that remains roughly constant throughout the stimulus period [26]. Even recent theoretical work that posits a strong feedback component for choice probabilities finds that this flat temporal choice probability profile reflects an early feedforward contribution to perceptual decisions [81, 251, 82]. Dynamics in V1 are less consistent with previous observations, and strongly suggest a feedback origin [154]. There was no choice probability in V1 early in the trial, but there was very little stimulus information in these neurons early in the trial as well. Thus, our data appear consistent with an interpretation whereby V2 neurons directly participate in supporting naturalistic texture discrimination, while both stimulus selectivity and choice-related activity in V1 arise via feedback connections from V2 or higher areas.

References

[1] John F. Ackermann & Michael S. Landy. “Statistical templates for visual search”. eng. In: *Journal of Vision* 14.3 (2014), p. 18.

[2] Daniel L. Adams, John R. Economides, Cristina M. Jocson & Jonathan C. Horton. “A biocompatible titanium headpost for stabilizing behaving monkeys”. eng. In: *Journal of Neurophysiology* 98.2 (Aug. 2007), pp. 993–1001.

[3] Daniel L. Adams, John R. Economides, Cristina M. Jocson, John M. Parker & Jonathan C. Horton. “A watertight acrylic-free titanium recording chamber for electrophysiology in behaving monkeys”. eng. In: *Journal of Neurophysiology* 106.3 (Sept. 2011), pp. 1581–1590.

[4] E. H. Adelson & J. R. Bergen. “Spatiotemporal energy models for the perception of motion”. eng. In: *Journal of the Optical Society of America. A, Optics and Image Science* 2.2 (Feb. 1985), pp. 284–299.

[5] Edward H. Adelson. “On seeing stuff: the perception of materials by humans and machines”. In: vol. 4299. 2001, pp. 1–12.

[6] Edward H. Adelson & James R. Bergen. “The Plenoptic Function and the Elements of Early Vision”. In: *Computational Models of Visual Processing*. MIT Press, 1991, pp. 3–20.

[7] Seyed-Reza Afraz, Roozbeh Kiani & Hossein Esteky. “Microstimulation of inferotemporal cortex influences face categorization”. eng. In: *Nature* 442.7103 (Aug. 2006), pp. 692–695.

[8] Henry J. Alitto, Bartlett D. Moore, Daniel L. Rathbun & W. Martin Usrey. "A comparison of visual responses in the lateral geniculate nucleus of alert and anaesthetized macaque monkeys". eng. In: *The Journal of Physiology* 589.Pt 1 (Jan. 2011), pp. 87–99.

[9] J. Allman, F. Miezin & E. McGuinness. "Stimulus specific responses from beyond the classical receptive field: neurophysiological mechanisms for local-global comparisons in visual neurons". eng. In: *Annual Review of Neuroscience* 8 (1985), pp. 407–430.

[10] Alessandra Angelucci, Jonathan B. Levitt, Emma J. S. Walton, Jean-Michel Hupe, Jean Bullier & Jennifer S. Lund. "Circuits for local and global signal integration in primary visual cortex". eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 22.19 (Oct. 2002), pp. 8633–8646.

[11] Akiyuki Anzai, Xinmiao Peng & David C. Van Essen. "Neurons in monkey visual area V2 encode combinations of orientations". eng. In: *Nature Neuroscience* 10.10 (Oct. 2007), pp. 1313–1321.

[12] F. Arcizet, C. Jouffrais & P. Girard. "Natural textures classification in area V4 of the macaque monkey". eng. In: *Experimental Brain Research* 189.1 (July 2008), pp. 109–120.

[13] Fred Attneave. "Some informational aspects of visual perception". In: *Psychological Review* 61.3 (1954), pp. 183–193.

[14] Tom Baden, Philipp Berens, Katrin Franke, Miroslav Román Rosón, Matthias Bethge & Thomas Euler. "The functional diversity of retinal ganglion cells in the mouse". eng. In: *Nature* 529.7586 (Jan. 2016), pp. 345–350.

[15] Wyeth Bair, James R. Cavanaugh & J. Anthony Movshon. "Time course and time-distance relationships for surround suppression in macaque V1 neurons". eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 23.20 (Aug. 2003), pp. 7690–7701.

[16] Benjamin J. Balas. "Texture synthesis and perception: using computational models to study texture representations in the human visual system". eng. In: *Vision Research* 46.3 (Feb. 2006), pp. 299–309.

[17] Benjamin Balas, Lisa Nakano & Ruth Rosenholtz. "A summary-statistic representation in peripheral vision explains visual crowding". eng. In: *Journal of Vision* 9.12 (2009), pp. 13.1–18.

[18] H. B. Barlow. "Why have multiple cortical areas?" In: *Vision Research* 26.1 (1986), pp. 81–90.

[19] H. B. Barlow & W. R. Levick. "The mechanism of directionally selective units in rabbit's retina". eng. In: *The Journal of Physiology* 178.3 (June 1965), pp. 477–504.

[20] HB Barlow. "Possible principles underlying the transformations of sensory messages". In: *Sensory Communication*. Ed. by WA Rosenblith. MIT Press, 1961, pp. 217–234.

[21] C. Blakemore & E. A. Tobin. "Lateral inhibition between orientation detectors in the cat's visual cortex". eng. In: *Experimental Brain Research* 15.4 (1972), pp. 439–440.

[22] Richard T. Born & David C. Bradley. "Structure and Function of Visual Area Mt". In: *Annual Review of Neuroscience* 28.1 (2005), pp. 157–189.

[23] C. E. Bredfeldt, J. C. A. Read & B. G. Cumming. "A quantitative explanation of responses to disparity-defined edges in macaque V2". eng. In: *Journal of Neurophysiology* 101.2 (Feb. 2009), pp. 701–713.

[24] Christine E. Bredfeldt & Bruce G. Cumming. "A simple account of cyclopean edge responses in macaque v2". eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 26.29 (July 2006), pp. 7581–7596.

[25] Kevin L. Briggman, Moritz Helmstaedter & Winfried Denk. "Wiring specificity in the direction-selectivity circuit of the retina". eng. In: *Nature* 471.7337 (Mar. 2011), pp. 183–188.

[26] K. H. Britten, W. T. Newsome, M. N. Shadlen, S. Celebrini & J. A. Movshon. "A relationship between behavioral choice and the visual responses of neurons in macaque MT". eng. In: *Visual Neuroscience* 13.1 (Feb. 1996), pp. 87–100.

[27] K. H. Britten, M. N. Shadlen, W. T. Newsome & J. A. Movshon. “The analysis of visual motion: a comparison of neuronal and psychophysical performance”. en. In: *The Journal of Neuroscience* 12.12 (Dec. 1992), pp. 4745–4765.

[28] Charles F. Cadieu, Ha Hong, Daniel L. K. Yamins, Nicolas Pinto, Diego Ardila, Ethan A. Solomon, Najib J. Majaj & James J. DiCarlo. “Deep neural networks rival the representation of primate IT cortex for core visual object recognition”. eng. In: *PLoS computational biology* 10.12 (Dec. 2014), e1003963.

[29] T. Caelli & B. Julesz. “On perceptual analyzers underlying visual texture discrimination: part I”. eng. In: *Biological Cybernetics* 28.3 (Feb. 1978), pp. 167–175.

[30] T. Caelli, B. Julesz & E. Gilbert. “On perceptual analyzers underlying visual texture discrimination: Part II”. eng. In: *Biological Cybernetics* 29.4 (June 1978), pp. 201–214.

[31] M. Carandini, D. J. Heeger & J. A. Movshon. “Linearity and normalization in simple cells of the macaque primary visual cortex”. eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 17.21 (Nov. 1997), pp. 8621–8644.

[32] Matteo Carandini, Jonathan B. Demb, Valerio Mante, David J. Tolhurst, Yang Dan, Bruno A. Olshausen, Jack L. Gallant & Nicole C. Rust. “Do we know what the early visual system does?” eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 25.46 (Nov. 2005), pp. 10577–10597.

[33] Matteo Carandini & David J. Heeger. “Normalization as a canonical neural computation”. eng. In: *Nature Reviews. Neuroscience* 13.1 (Jan. 2012), pp. 51–62.

[34] James R. Cavanaugh, Wyeth Bair & J. Anthony Movshon. “Nature and interaction of signals from the receptive field center and surround in macaque V1 neurons”. eng. In: *Journal of Neurophysiology* 88.5 (Nov. 2002), pp. 2530–2546.

[35] James R. Cavanaugh, Wyeth Bair & J. Anthony Movshon. “Selectivity and spatial distribution of signals from the receptive field surround in macaque V1 neurons”. eng. In: *Journal of Neurophysiology* 88.5 (Nov. 2002), pp. 2547–2556.

[36] S. Celebrini & W. T. Newsome. “Neuronal and psychophysical sensitivity to motion signals in extrastriate area MST of the macaque monkey”. eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 14.7 (July 1994), pp. 4109–4124.

[37] Xiaodong Chen, Feng Han, Mu-Ming Poo & Yang Dan. “Excitatory and suppressive receptive field subunits in awake monkey primary visual cortex (V1)”. eng. In: *Proceedings of the National Academy of Sciences of the United States of America* 104.48 (Nov. 2007), pp. 19120–19125.

[38] Ruben Coen-Cagli, Adam Kohn & Odelia Schwartz. “Flexible gating of contextual influences in natural vision”. en. In: *Nature Neuroscience* advance online publication (Oct. 2015).

[39] Marlene R. Cohen & William T. Newsome. “Estimates of the contribution of single neurons to perception depend on timescale and noise correlation”. eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 29.20 (May 2009), pp. 6635–6648.

[40] Edward Craft, Hartmut Schütze, Ernst Niebur & Rüdiger von der Heydt. “A neural model of figure-ground organization”. eng. In: *Journal of Neurophysiology* 97.6 (June 2007), pp. 4310–4326.

[41] A. P. Dawid & A. M. Skene. “Maximum Likelihood Estimation of Observer Error-Rates Using the EM Algorithm”. In: *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 28.1 (1979), pp. 20–28.

[42] R. L. De Valois, D. G. Albrecht & L. G. Thorell. “Spatial frequency selectivity of cells in macaque visual cortex”. eng. In: *Vision Research* 22.5 (1982), pp. 545–559.

[43] G. C. DeAngelis, R. D. Freeman & I. Ohzawa. “Length and width tuning of neurons in the cat’s primary visual cortex”. eng. In: *Journal of Neurophysiology* 71.1 (Jan. 1994), pp. 347–374.

[44] Edgar A. DeYoe & David C. Van Essen. “Segregation of efferent connections and receptive field properties in visual area V2 of the macaque”. en. In: *Nature* 317.6032 (Sept. 1985), pp. 58–61.

[45] James J. DiCarlo & David D. Cox. “Untangling invariant object recognition”. eng. In: *Trends in Cognitive Sciences* 11.8 (Aug. 2007), pp. 333–341.

[46] James J. DiCarlo, Davide Zoccolan & Nicole C. Rust. “How does the brain solve visual object recognition?” In: *Neuron* 73.3 (Feb. 2012), pp. 415–434.

[47] C. Enroth-Cugell & J. G. Robson. “The contrast sensitivity of retinal ganglion cells of the cat”. eng. In: *The Journal of Physiology* 187.3 (Dec. 1966), pp. 517–552.

[48] Fang Fang, Huseyin Boyaci & Daniel Kersten. “Border ownership selectivity in human early visual cortex and its modulation by attention”. eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 29.2 (Jan. 2009), pp. 460–465.

[49] Frederick Federer, Jennifer M. Ichida, Janelle Jeffs, Ingo Schiessl, Niall McLoughlin & Alessandra Angelucci. “Four projection streams from primate V1 to the cytochrome oxidase stripes of V2”. eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 29.49 (Dec. 2009), pp. 15455–15471.

[50] Frederick Federer, Delaney Williams, Jennifer M. Ichida, Sam Merlin & Alessandra Angelucci. “Two projection streams from macaque V1 to the pale cytochrome oxidase stripes of V2”. eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 33.28 (July 2013), pp. 11530–11539.

[51] D. J. Felleman & D. C. Van Essen. “Receptive field properties of neurons in area V3 of macaque monkey extrastriate cortex”. eng. In: *Journal of Neurophysiology* 57.4 (Apr. 1987), pp. 889–920.

[52] Daniel J. Felleman & David C. Van Essen. “Distributed Hierarchical Processing in the Primate Cerebral Cortex”. en. In: *Cerebral Cortex* 1.1 (Jan. 1991), pp. 1–47.

[53] Gidon Felsen & Yang Dan. “A natural approach to studying vision”. eng. In: *Nature Neuroscience* 8.12 (Dec. 2005), pp. 1643–1646.

[54] Gidon Felsen, Jon Touryan, Feng Han & Yang Dan. “Cortical sensitivity to visual features in natural scenes”. eng. In: *PLoS biology* 3.10 (Oct. 2005), e342.

[55] D. Ferster, S. Chung & H. Wheat. “Orientation selectivity of thalamic input to simple cells of cat visual cortex”. eng. In: *Nature* 380.6571 (Mar. 1996), pp. 249–252.

[56] G. D. Field & E. J. Chichilnisky. “Information Processing in the Primate Retina: Circuitry and Coding”. In: *Annual Review of Neuroscience* 30.1 (2007), pp. 1–30.

[57] K. H. Foster, J. P. Gaska, M. Nagler & D. A. Pollen. “Spatial and temporal frequency selectivity of neurones in visual cortical areas V1 and V2 of the macaque monkey”. eng. In: *The Journal of Physiology* 365 (Aug. 1985), pp. 331–363.

[58] Jeremy Freeman. “Computation and representation in the primate visual system”. PhD thesis. Center for Neural Science, New York University, 2012.

[59] Jeremy Freeman, Greg D. Field, Peter H. Li, Martin Greschner, Deborah E. Gunning, Keith Mathieson, Alexander Sher, Alan M. Litke, Liam Paninski, Eero P. Simoncelli & E. J. Chichilnisky. “Mapping nonlinear receptive field structure in primate retina at single cone resolution”. en. In: *eLife* 4 (Oct. 2015), e05241.

[60] Jeremy Freeman & Eero P. Simoncelli. “Metamers of the ventral stream”. eng. In: *Nature Neuroscience* 14.9 (Sept. 2011), pp. 1195–1201.

[61] Jeremy Freeman, Corey M. Ziemba, David J. Heeger, Eero P. Simoncelli & J. Anthony Movshon. “A functional and perceptual signature of the second visual area in primates”. en. In: *Nature Neuroscience* 16.7 (July 2013), pp. 974–981.

[62] Murray F. Freeman, John W. Tukey, et al. “Transformations Related to the Angular and the Square Root”. In: *The Annals of Mathematical Statistics* 21.4 (1950), pp. 607–611.

[63] Justin L. Gardner, Elisha P. Merriam, J. Anthony Movshon & David J. Heeger. “Maps of visual space in human occipital cortex are retinotopic, not spatiotopic”. eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 28.15 (Apr. 2008), pp. 3988–3999.

[64] R. Gattass, C. G. Gross & J. H. Sandell. “Visual topography of V2 in the macaque”. eng. In: *The Journal of Comparative Neurology* 201.4 (Oct. 1981), pp. 519–539.

[65] R. Gattass, A. P. Sousa & C. G. Gross. “Visuotopic organization and extent of V3 and V4 of the macaque”. eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 8.6 (June 1988), pp. 1831–1845.

[66] R. Gattass, A. P. Sousa, M. Mishkin & L. G. Ungerleider. “Cortical projections of area V2 in the macaque”. eng. In: *Cerebral Cortex (New York, N.Y.: 1991)* 7.2 (Mar. 1997), pp. 110–129.

[67] Leon A. Gatys, Alexander S. Ecker & Matthias Bethge. “Texture Synthesis Using Convolutional Neural Networks”. In: *arXiv:1505.07376 [cs, q-bio]* (May 2015).

[68] K. R. Gegenfurtner, D. C. Kiper & S. B. Fenstemaker. “Processing of color, form, and motion in macaque area V2”. eng. In: *Visual Neuroscience* 13.1 (Feb. 1996), pp. 161–172.

[69] W. S. Geisler, J. S. Perry, B. J. Super & D. P. Gallogly. “Edge co-occurrence in natural images predicts contour grouping performance”. eng. In: *Vision Research* 41.6 (Mar. 2001), pp. 711–724.

[70] C. D. Gilbert, A. Das, M. Ito, M. Kapadia & G. Westheimer. “Spatial integration and cortical dynamics”. eng. In: *Proceedings of the National Academy of Sciences of the United States of America* 93.2 (Jan. 1996), pp. 615–622.

[71] C. D. Gilbert & T. N. Wiesel. “The influence of contextual stimuli on the orientation selectivity of cells in primary visual cortex of the cat”. eng. In: *Vision Research* 30.11 (1990), pp. 1689–1701.

[72] Lindsey L. Glickfeld, Mark L. Andermann, Vincent Bonin & R. Clay Reid. "Cortico-cortical projections in mouse visual cortex are functionally target specific". eng. In: *Nature Neuroscience* 16.2 (Feb. 2013), pp. 219–226.

[73] Ian J. Goodfellow, Jonathon Shlens & Christian Szegedy. "Explaining and Harnessing Adversarial Examples". In: *arXiv:1412.6572 [cs, stat]* (Dec. 2014).

[74] Robbe L. T. Goris, J. Anthony Movshon & Eero P. Simoncelli. "Partitioning neuronal variability". eng. In: *Nature Neuroscience* 17.6 (June 2014), pp. 858–865.

[75] Robbe L. T. Goris, Eero P. Simoncelli & J. Anthony Movshon. "Origin and Function of Tuning Diversity in Macaque Visual Cortex". eng. In: *Neuron* 88.4 (Nov. 2015), pp. 819–831.

[76] John A. Greenwood, Peter J. Bex & Steven C. Dakin. "Positional averaging explains crowding with letter-like stimuli". eng. In: *Proceedings of the National Academy of Sciences of the United States of America* 106.31 (Aug. 2009), pp. 13130–13135.

[77] Ulrike Grömping. "Estimators of Relative Importance in Linear Regression Based on Variance Decomposition". In: *The American Statistician* 61.2 (May 2007), pp. 139–147.

[78] C. G. Gross, C. E. Rocha-Miranda & D. B. Bender. "Visual properties of neurons in inferotemporal cortex of the Macaque." en. In: *Journal of Neurophysiology* 35.1 (Jan. 1972), pp. 96–111.

[79] Yong Gu, Dora E. Angelaki & Gregory C. DeAngelis. "Neural correlates of multisensory cue integration in macaque MSTd". en. In: *Nature Neuroscience* 11.10 (Oct. 2008), pp. 1201–1210.

[80] Yong Gu, Gregory C. DeAngelis & Dora E. Angelaki. "A functional link between area MSTd and heading perception based on vestibular signals". eng. In: *Nature Neuroscience* 10.8 (Aug. 2007), pp. 1038–1047.

[81] Ralf M. Haefner, Pietro Berkes & József Fiser. "The implications of perception as probabilistic inference for correlated neural variability during behavior". In: *arXiv:1409.0257 [q-bio]* (Aug. 2014).

[82] Ralf M. Haefner, Pietro Berkes & József Fiser. “Perceptual Decision-Making as Probabilistic Inference by Neural Sampling”. English. In: *Neuron* (Apr. 2016).

[83] Luke E. Hallum & J. Anthony Movshon. “Surround suppression supports second-order feature encoding by macaque V1 and V2 neurons”. In: *Vision Research*. The Function of Contextual Modulation 104 (Nov. 2014), pp. 24–35.

[84] James H. Hedges, Yevgeniya Gartshteyn, Adam Kohn, Nicole C. Rust, Michael N. Shadlen, William T. Newsome & J. Anthony Movshon. “Dissociation of neuronal and psychophysical responses to local and global motion”. eng. In: *Current biology: CB* 21.23 (Dec. 2011), pp. 2023–2028.

[85] D. J. Heeger. “Normalization of cell responses in cat striate cortex”. eng. In: *Visual Neuroscience* 9.2 (Aug. 1992), pp. 181–197.

[86] D. J. Heeger, G. M. Boynton, J. B. Demb, E. Seidemann & W. T. Newsome. “Motion opponency in visual cortex”. eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 19.16 (Aug. 1999), pp. 7162–7174.

[87] D. J. Heeger, E. P. Simoncelli & J. A. Movshon. “Computational models of cortical visual processing”. eng. In: *Proceedings of the National Academy of Sciences of the United States of America* 93.2 (Jan. 1996), pp. 623–627.

[88] David J. Heeger & James R. Bergen. “Pyramid-based Texture Analysis/Synthesis”. In: *Proceedings of the 22Nd Annual Conference on Computer Graphics and Interactive Techniques*. SIGGRAPH ’95. New York, NY, USA: ACM, 1995, pp. 229–238.

[89] J. Hegdé & D. C. Van Essen. “Selectivity for complex shapes in primate visual area V2”. eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 20.5 (Mar. 2000), RC61.

[90] Jay Hegdé & David C. Van Essen. “Temporal dynamics of shape analysis in macaque visual area V2”. eng. In: *Journal of Neurophysiology* 92.5 (Nov. 2004), pp. 3030–3042.

[91] Jay Hegdé & David C. Van Essen. “A comparative study of shape representation in macaque visual areas v2 and v4”. eng. In: *Cerebral Cortex (New York, N.Y.: 1991)* 17.5 (May 2007), pp. 1100–1116.

[92] O. J. Henaff, R. L. T. Goris & E. P. Simoncelli. “Perceptual evaluation of artificial visual recognition systems using geodesics”. In: *Computational and Systems Neuroscience (CoSyNe)*. 2015.

[93] Christopher A. Henry, Siddhartha Joshi, Dajun Xing, Robert M. Shapley & Michael J. Hawken. “Functional characterization of the extraclassical receptive field in macaque V1: contrast, orientation, and temporal dynamics”. eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 33.14 (Apr. 2013), pp. 6230–6242.

[94] Rüdiger von der Heydt. “Figure–ground organization and the emergence of proto-objects in the visual cortex”. In: *Perception Science* (2015), p. 1695.

[95] J. M. Hillis, M. O. Ernst, M. S. Banks & M. S. Landy. “Combining sensory information: mandatory fusion within, but not between, senses”. eng. In: *Science (New York, N.Y.)* 298.5598 (Nov. 2002), pp. 1627–1630.

[96] Geoffrey E. Hinton & Sam T. Roweis. “Stochastic neighbor embedding”. In: *Advances in neural information processing systems*. 2002, pp. 833–840.

[97] S. Hochstein & R. M. Shapley. “Linear and nonlinear spatial subunits in Y cat retinal ganglion cells”. eng. In: *The Journal of Physiology* 262.2 (Nov. 1976), pp. 265–284.

[98] Jonathan C Horton & Daniel L Adams. “The cortical column: a structure without a function”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 360.1456 (Apr. 2005), pp. 837–862.

[99] D. H. Hubel & M. S. Livingstone. “Complex-unoriented cells in a subregion of primate area 18”. eng. In: *Nature* 315.6017 (May 1985), pp. 325–327.

[100] D. H. Hubel & M. S. Livingstone. “Segregation of form, color, and stereopsis in primate area 18”. eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 7.11 (Nov. 1987), pp. 3378–3415.

[101] D. H. Hubel & T. N. Wiesel. “Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex”. eng. In: *The Journal of Physiology* 160 (Jan. 1962), pp. 106–154.

[102] D. H. Hubel & T. N. Wiesel. “Receptive fields and functional architecture of monkey striate cortex”. eng. In: *The Journal of Physiology* 195.1 (Mar. 1968), pp. 215–243.

[103] David H. Hubel & Torsten N. Wiesel. “Receptive Fields and Functional Architecture in Two Nonstriate Visual Areas (18 and 19) of the Cat”. en. In: *Journal of Neurophysiology* 28.2 (Mar. 1965), pp. 229–289.

[104] Chou P. Hung, Gabriel Kreiman, Tomaso Poggio & James J. DiCarlo. “Fast readout of object identity from macaque inferior temporal cortex”. eng. In: *Science (New York, N.Y.)* 310.5749 (Nov. 2005), pp. 863–866.

[105] Minami Ito & Naokazu Goda. “Mechanisms underlying the representation of angles embedded within contour stimuli in area V2 of macaque monkeys”. eng. In: *The European Journal of Neuroscience* 33.1 (Jan. 2011), pp. 130–142.

[106] Minami Ito & Hidehiko Komatsu. “Representation of angles embedded within contour stimuli in area V2 of macaque monkeys”. eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 24.13 (Mar. 2004), pp. 3313–3324.

[107] Mehrdad Jazayeri, Pascal Wallisch & J. Anthony Movshon. “Dynamics of Macaque MT Cell Responses to Grating Triplets”. en. In: *The Journal of Neuroscience* 32.24 (June 2012), pp. 8242–8253.

[108] B. Julesz. “Visual Pattern Discrimination”. In: *IRE Transactions on Information Theory* 8.2 (Feb. 1962), pp. 84–92.

[109] B. Julesz. “Textons, the elements of texture perception, and their interactions”. eng. In: *Nature* 290.5802 (Mar. 1981), pp. 91–97.

[110] B. Julesz, E. N. Gilbert & J. D. Victor. “Visual discrimination of textures with identical third-order statistics”. eng. In: *Biological Cybernetics* 31.3 (Dec. 1978), pp. 137–140.

[111] Yan Karklin & Michael S. Lewicki. “Emergence of complex cell properties by learning to generalize in natural scenes”. eng. In: *Nature* 457.7225 (Jan. 2009), pp. 83–86.

[112] Seyed-Mahdi Khaligh-Razavi & Nikolaus Kriegeskorte. “Deep supervised, but not unsupervised, models may explain IT cortical representation”. eng. In: *PLoS computational biology* 10.11 (Nov. 2014), e1003915.

[113] Roozbeh Kiani, Hossein Esteky, Koorosh Mirpour & Keiji Tanaka. “Object category structure in response patterns of neuronal population in monkey inferior temporal cortex”. eng. In: *Journal of Neurophysiology* 97.6 (June 2007), pp. 4296–4309.

[114] D. C. Kiper, S. B. Fenstemaker & K. R. Gegenfurtner. “Chromatic properties of neurons in macaque area V2”. eng. In: *Visual Neuroscience* 14.6 (Dec. 1997), pp. 1061–1072.

[115] J. J. Knierim & D. C. van Essen. “Neuronal responses to static texture patterns in area V1 of the alert macaque monkey”. eng. In: *Journal of Neurophysiology* 67.4 (Apr. 1992), pp. 961–980.

[116] Nikolaus Kriegeskorte. “Deep neural networks: a new framework for modelling biological vision and brain information processing”. en. In: *bioRxiv* (Oct. 2015), p. 029876.

[117] Nikolaus Kriegeskorte, Marieke Mur & Peter Bandettini. “Representational similarity analysis - connecting the branches of systems neuroscience”. eng. In: *Frontiers in Systems Neuroscience* 2 (2008), p. 4.

[118] Nikolaus Kriegeskorte, Marieke Mur, Douglas A. Ruff, Roozbeh Kiani, Jerzy Bodurka, Hossein Esteky, Keiji Tanaka & Peter A. Bandettini. “Matching categorical object representations in inferior temporal cortex of man and monkey”. eng. In: *Neuron* 60.6 (Dec. 2008), pp. 1126–1141.

[119] Jonas Larsson & David J. Heeger. “Two retinotopic visual areas in human lateral occipital cortex”. eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 26.51 (Dec. 2006), pp. 13128–13142.

[120] T. S. Lee & M. Nguyen. “Dynamics of subjective contour formation in the early visual cortex”. eng. In: *Proceedings of the National Academy of Sciences of the United States of America* 98.4 (Feb. 2001), pp. 1907–1911.

[121] P. Lennie. “Single units and visual cortical organization”. eng. In: *Perception* 27.8 (1998), pp. 889–935.

[122] Jerome Y. Lettvin. “On Seeing Sidelong”. en. In: *The Sciences* 16.4 (July 1976), pp. 10–20.

[123] J. B. Levitt, D. C. Kiper & J. A. Movshon. “Receptive fields and functional architecture of macaque V2”. eng. In: *Journal of Neurophysiology* 71.6 (June 1994), pp. 2517–2542.

[124] J. B. Levitt & J. S. Lund. “Contrast dependence of contextual effects in primate visual cortex”. eng. In: *Nature* 387.6628 (May 1997), pp. 73–76.

[125] Guangxing Li, Zhimo Yao, Zhengchun Wang, Nini Yuan, Vargha Talebi, Jiabo Tan, Yongchang Wang, Yifeng Zhou & Curtis L. Baker. “Form-cue invariant second-order neuronal responses to contrast modulation in primate area V2”. eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 34.36 (Sept. 2014), pp. 12081–12092.

[126] Wu Li & Charles D. Gilbert. “Global contour saliency and local colinear interactions”. eng. In: *Journal of Neurophysiology* 88.5 (Nov. 2002), pp. 2846–2856.

[127] Wu Li, Valentin Piëch & Charles D. Gilbert. “Contour saliency in primary visual cortex”. eng. In: *Neuron* 50.6 (June 2006), pp. 951–962.

[128] Wu Li, Valentin Piëch & Charles D. Gilbert. “Learning to link visual contours”. eng. In: *Neuron* 57.3 (Feb. 2008), pp. 442–451.

[129] Lu Liu, Liang She, Ming Chen, Tianyi Liu, Haidong D. Lu, Yang Dan & Mu-Ming Poo. “Spatial structure of neuronal receptive field in awake monkey secondary visual cortex (V2)”. eng. In: *Proceedings of the National Academy of Sciences of the United States of America* 113.7 (Feb. 2016), pp. 1913–1918.

[130] Timm Lochmann, Timothy J. Blanche & Daniel A. Butts. “Construction of direction selectivity through local energy computations in primary visual cortex”. eng. In: *PloS One* 8.3 (2013), e58666.

[131] Haidong D. Lu, Gang Chen, Hisashi Tanigawa & Anna W. Roe. “A motion direction map in macaque V2”. eng. In: *Neuron* 68.5 (Dec. 2010), pp. 1002–1013.

[132] L. E. Mahon & R. L. De Valois. “Cartesian and non-Cartesian responses in LGN, V1, and V2 cells”. eng. In: *Visual Neuroscience* 18.6 (Dec. 2001), pp. 973–981.

[133] R. Malach, J. B. Reppas, R. R. Benson, K. K. Kwong, H. Jiang, W. A. Kennedy, P. J. Ledden, T. J. Brady, B. R. Rosen & R. B. Tootell. “Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex”. eng. In: *Proceedings of the National Academy of Sciences of the United States of America* 92.18 (Aug. 1995), pp. 8135–8139.

[134] Valerio Mante, Vincent Bonin & Matteo Carandini. “Functional Mechanisms Shaping Lateral Geniculate Responses to Artificial and Natural Stimuli”. In: *Neuron* 58.4 (May 2008), pp. 625–638.

[135] Richard H. Masland. “The neuronal organization of the retina”. eng. In: *Neuron* 76.2 (Oct. 2012), pp. 266–280.

[136] Josh H. McDermott, Michael Schemitsch & Eero P. Simoncelli. “Summary statistics in auditory perception”. eng. In: *Nature Neuroscience* 16.4 (Apr. 2013), pp. 493–498.

[137] Josh H. McDermott & Eero P. Simoncelli. “Sound texture perception via statistics of the auditory periphery: evidence from sound synthesis”. eng. In: *Neuron* 71.5 (Sept. 2011), pp. 926–940.

[138] W. H. Merigan, T. A. Nealey & J. H. Maunsell. “Visual effects of lesions of cortical area V2 in macaques”. en. In: *The Journal of Neuroscience* 13.7 (July 1993), pp. 3180–3191.

[139] Kenneth D. Miller. “Canonical computations of cerebral cortex”. ENG. In: *Current Opinion in Neurobiology* 37 (Feb. 2016), pp. 75–84.

[140] J. A. Movshon & W. T. Newsome. “Visual response properties of striate cortical neurons projecting to area MT in macaque monkeys”. eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 16.23 (Dec. 1996), pp. 7733–7741.

[141] J. A. Movshon, I. D. Thompson & D. J. Tolhurst. “Receptive field organization of complex cells in the cat’s striate cortex”. eng. In: *The Journal of Physiology* 283 (Oct. 1978), pp. 79–99.

[142] J. A. Movshon, I. D. Thompson & D. J. Tolhurst. “Spatial and temporal contrast sensitivity of neurones in areas 17 and 18 of the cat’s visual cortex”. eng. In: *The Journal of Physiology* 283 (Oct. 1978), pp. 101–120.

[143] J. A. Movshon, I. D. Thompson & D. J. Tolhurst. “Spatial summation in the receptive fields of simple cells in the cat’s striate cortex”. eng. In: *The Journal of Physiology* 283 (Oct. 1978), pp. 53–77.

[144] J. Anthony Movshon, E. H. Adelson, M. S. Gizzi & William T. Newsome. “The analysis of moving visual patterns”. English (US). In: 1985.

[145] H. Nakamura, R. Gattass, R. Desimone & L. G. Ungerleider. “The modular organization of projections from areas V1 and V2 to areas V4 and TEO in macaques”. eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 13.9 (Sept. 1993), pp. 3681–3691.

[146] Jonathan J. Nassi, Stephen G. Lomber & Richard T. Born. “Corticocortical feedback contributes to surround suppression in V1 of the alert primate”. eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 33.19 (May 2013), pp. 8504–8517.

[147] O. Nestares & D. J. Heeger. “Robust multiresolution alignment of MRI brain volumes”. eng. In: *Magnetic Resonance in Medicine* 43.5 (May 2000), pp. 705–715.

[148] W. T. Newsome & E. B. Paré. “A selective impairment of motion perception following lesions of the middle temporal visual area (MT)”. eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 8.6 (June 1988), pp. 2201–2211.

[149] William T. Newsome, Kenneth H. Britten & J. Anthony Movshon. “Neuronal correlates of a perceptual decision”. en. In: *Nature* 341.6237 (Sept. 1989), pp. 52–54.

[150] A. Nguyen, J. Yosinski & J. Clune. “Deep neural networks are easily fooled: High confidence predictions for unrecognizable images”. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2015, pp. 427–436.

[151] Hendrikje Nienborg, Marlene R. Cohen & Bruce G. Cumming. “Decision-related activity in sensory neurons: correlations among neurons and with behavior”. eng. In: *Annual Review of Neuroscience* 35 (2012), pp. 463–483.

[152] Hendrikje Nienborg & Bruce G. Cumming. “Macaque V2 Neurons, But Not V1 Neurons, Show Choice-Related Activity”. en. In: *The Journal of Neuroscience* 26.37 (Sept. 2006), pp. 9567–9578.

[153] Hendrikje Nienborg & Bruce G. Cumming. “Psychophysically measured task strategy for disparity discrimination is reflected in V2 neurons”. eng. In: *Nature Neuroscience* 10.12 (Dec. 2007), pp. 1608–1614.

[154] Hendrikje Nienborg & Bruce G. Cumming. “Decision-related activity in sensory neurons reflects more than a neuron’s causal effect”. eng. In: *Nature* 459.7243 (May 2009), pp. 89–92.

[155] Hendrikje Nienborg & Bruce G. Cumming. “Decision-related activity in sensory neurons may depend on the columnar architecture of cerebral cortex”. eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 34.10 (Mar. 2014), pp. 3579–3585.

[156] Shinji Nishimoto & Jack L. Gallant. “A three-dimensional spatiotemporal receptive field model explains responses of area MT neurons to naturalistic movies”. eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 31.41 (Oct. 2011), pp. 14551–14564.

[157] Shinji Nishimoto, Tsugitaka Ishida & Izumi Ohzawa. “Receptive field properties of neurons in the early visual cortex revealed by local spectral reverse correlation”. eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 26.12 (Mar. 2006), pp. 3269–3280.

[158] S. Ogawa, T. M. Lee, A. R. Kay & D. W. Tank. “Brain magnetic resonance imaging with contrast dependent on blood oxygenation”. eng. In: *Proceedings of the National Academy of Sciences of the United States of America* 87.24 (Dec. 1990), pp. 9868–9872.

[159] Philip O’Herron & Rüdiger von der Heydt. “Short-term memory for figure-ground organization in the visual cortex”. eng. In: *Neuron* 61.5 (Mar. 2009), pp. 801–809.

[160] Philip O’Herron & Rüdiger von der Heydt. “Remapping of border ownership in the visual cortex”. eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 33.5 (Jan. 2013), pp. 1964–1974.

[161] Gouki Okazawa, Satohiro Tajima & Hidehiko Komatsu. “Gradual development of visual texture-selective properties between macaque areas V2 and V4”. In: *Submitted* ().

[162] Gouki Okazawa, Satohiro Tajima & Hidehiko Komatsu. “Image statistics underlying natural texture selectivity of neurons in macaque V4”. eng. In: *Proceedings of the National Academy of Sciences of the United States of America* 112.4 (Jan. 2015), E351–360.

[163] G. A. Orban, H. Kennedy & J. Bullier. “Velocity sensitivity and direction selectivity of neurons in areas V1 and V2 of the monkey: influence of eccentricity”. eng. In: *Journal of Neurophysiology* 56.2 (Aug. 1986), pp. 462–480.

[164] Gabriele Paolacci, Jesse Chandler & Panagiotis G. Ipeirotis. *Running Experiments on Amazon Mechanical Turk*. SSRN Scholarly Paper ID 1626226. Rochester, NY: Social Science Research Network, June 2010.

[165] A. J. Parker & W. T. Newsome. “SENSE AND THE SINGLE NEURON: Probing the Physiology of Perception”. In: *Annual Review of Neuroscience* 21.1 (1998), pp. 227–277.

[166] L. Parkes, J. Lund, A. Angelucci, J. A. Solomon & M. Morgan. “Compulsory averaging of crowded orientation signals in human vision”. eng. In: *Nature Neuroscience* 4.7 (July 2001), pp. 739–744.

[167] Denis G. Pelli, Melanie Palomares & Najib J. Majaj. “Crowding is unlike ordinary masking: distinguishing feature integration from detection”. eng. In: *Journal of Vision* 4.12 (Dec. 2004), pp. 1136–1169.

[168] Denis G. Pelli & Katharine A. Tillman. “The uncrowded window of object recognition”. en. In: *Nature Neuroscience* 11.10 (Sept. 2008), pp. 1129–1135.

[169] E. Peterhans & R. von der Heydt. “Mechanisms of contour perception in monkey visual cortex. II. Contours bridging gaps”. en. In: *The Journal of Neuroscience* 9.5 (May 1989), pp. 1749–1763.

[170] Xaq Pitkow, Sheng Liu, Dora E. Angelaki, Gregory C. DeAngelis & Alexandre Pouget. “How Can Single Sensory Neurons Predict Behavior?” English. In: *Neuron* 87.2 (July 2015), pp. 411–423.

[171] Carlos R. Ponce, Stephen G. Lomber & Richard T. Born. “Integrating motion and depth via parallel pathways”. eng. In: *Nature Neuroscience* 11.2 (Feb. 2008), pp. 216–223.

[172] Javier Portilla & Eero P. Simoncelli. “A Parametric Texture Model Based on Joint Statistics of Complex Wavelet Coefficients”. en. In: *International Journal of Computer Vision* 40.1 (Oct. 2000), pp. 49–70.

[173] Nicholas J. Priebe & David Ferster. “Mechanisms of Neuronal Computation in Mammalian Visual Cortex”. In: *Neuron* 75.2 (July 2012), pp. 194–208.

[174] S. J. Prince, A. D. Pointon, B. G. Cumming & A. J. Parker. “The precision of single neuron responses in cortical area V1 during stereoscopic depth judgments”. eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 20.9 (May 2000), pp. 3387–3400.

[175] Gopathy Purushothaman & David C. Bradley. “Neural population code for fine perceptual decisions in area MT”. en. In: *Nature Neuroscience* 8.1 (Jan. 2005), pp. 99–106.

[176] Fangtu T. Qiu, Tadashi Sugihara & Rüdiger von der Heydt. “Figure-ground mechanisms provide structure for selective attention”. eng. In: *Nature Neuroscience* 10.11 (Nov. 2007), pp. 1492–1499.

[177] D. Ress, B. T. Backus & D. J. Heeger. “Activity in primary visual cortex predicts performance in a visual detection task”. eng. In: *Nature Neuroscience* 3.9 (Sept. 2000), pp. 940–945.

[178] Fred Rieke & Michael E. Rudd. “The Challenges Natural Images Pose for Visual Adaptation”. In: *Neuron* 64.5 (Dec. 2009), pp. 605–616.

[179] Maximilian Riesenhuber & Tomaso Poggio. “Hierarchical models of object recognition in cortex”. en. In: *Nature Neuroscience* 2.11 (Nov. 1999), pp. 1019–1025.

[180] Anna W. Roe, Leonardo Chelazzi, Charles E. Connor, Bevil R. Conway, Ichiro Fujita, Jack L. Gallant, Haidong Lu & Wim Vanduffel. “Toward a unified theory of visual area V4”. eng. In: *Neuron* 74.1 (Apr. 2012), pp. 12–29.

[181] Marcello G. P Rosa & Leah A Krubitzer. “The evolution of visual cortex: where is V2?” In: *Trends in Neurosciences* 22.6 (June 1999), pp. 242–248.

[182] null Ruderman & null Bialek. “Statistics of natural images: Scaling in the woods”. ENG. In: *Physical Review Letters* 73.6 (Aug. 1994), pp. 814–817.

[183] Nicole C. Rust & James J. DiCarlo. “Balanced increases in selectivity and tolerance produce constant sparseness along the ventral visual stream”. eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 32.30 (July 2012), pp. 10170–10182.

[184] Nicole C. Rust & James J. Dicarlo. “Selectivity and tolerance ("invariance") both increase as visual information propagates from cortical area V4 to IT”. eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 30.39 (Sept. 2010), pp. 12978–12995.

[185] Nicole C. Rust, Valerio Mante, Eero P. Simoncelli & J. Anthony Movshon. “How MT cells analyze the motion of visual patterns”. eng. In: *Nature Neuroscience* 9.11 (Nov. 2006), pp. 1421–1431.

[186] Nicole C. Rust & J. Anthony Movshon. “In praise of artifice”. eng. In: *Nature Neuroscience* 8.12 (Dec. 2005), pp. 1647–1650.

[187] Nicole C. Rust, Odelia Schwartz, J. Anthony Movshon & Eero P. Simoncelli. "Spatiotemporal elements of macaque v1 receptive fields". eng. In: *Neuron* 46.6 (June 2005), pp. 945–956.

[188] C. D. Salzman, K. H. Britten & W. T. Newsome. "Cortical microstimulation influences perceptual judgements of motion direction". eng. In: *Nature* 346.6280 (July 1990), pp. 174–177.

[189] C. D. Salzman, C. M. Murasugi, K. H. Britten & W. T. Newsome. "Microstimulation in visual area MT: effects on direction discrimination performance". eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 12.6 (June 1992), pp. 2331–2355.

[190] Peter H. Schiller & Joseph G. Malpeli. "The effect of striate cortex cooling on area 18 cells in the monkey". In: *Brain Research* 126.2 (May 1977), pp. 366–369.

[191] Anita M. Schmid, Keith P. Purpura & Jonathan D. Victor. "Responses to orientation discontinuities in V1 and V2: physiological dissociations and functional implications". eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 34.10 (Mar. 2014), pp. 3559–3578.

[192] Benjamin Scholl, Andrew Y. Y. Tan, Joseph Corey & Nicholas J. Priebe. "Emergence of orientation selectivity in the Mammalian visual pathway". eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 33.26 (June 2013), pp. 10616–10624.

[193] O. Schwartz & E. P. Simoncelli. "Natural signal statistics and sensory gain control". eng. In: *Nature Neuroscience* 4.8 (Aug. 2001), pp. 819–825.

[194] M. N. Shadlen, K. H. Britten, W. T. Newsome & J. A. Movshon. "A computational analysis of the relationship between neuronal and behavioral responses to visual motion". eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 16.4 (Feb. 1996), pp. 1486–1510.

[195] Yasmine El-Shamayleh, Romesh D. Kumbhani, Neel T. Dhruv & J. Anthony Movshon. "Visual response properties of V1 neurons projecting to V2 in

macaque". eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 33.42 (Oct. 2013), pp. 16594–16605.

[196] Yasmine El-Shamayleh & J. Anthony Movshon. "Neuronal responses to texture-defined form in macaque visual area V2". eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 31.23 (June 2011), pp. 8543–8555.

[197] Robert Shapley, Michael Hawken & Dajun Xing. "The dynamics of visual responses in the primary visual cortex". eng. In: *Progress in Brain Research* 165 (2007), pp. 21–32.

[198] Hiroshi M. Shiozaki, Seiji Tanabe, Takahiro Doi & Ichiro Fujita. "Neural Activity in Cortical Area V4 Underlies Fine Disparity Discrimination". en. In: *The Journal of Neuroscience* 32.11 (Mar. 2012), pp. 3830–3841.

[199] S. Shipp & S. Zeki. "Segregation of pathways leading from area V2 to areas V4 and V5 of macaque monkey visual cortex". en. In: *Nature* 315.6017 (May 1985), pp. 322–324.

[200] Stewart Shipp & Semir Zeki. "The functional organization of area V2, I: specialization across stripes and layers". eng. In: *Visual Neuroscience* 19.2 (Apr. 2002), pp. 187–210.

[201] S. Shushruth, Jennifer M. Ichida, Jonathan B. Levitt & Alessandra Angelucci. "Comparison of spatial summation properties of neurons in macaque V1 and V2". eng. In: *Journal of Neurophysiology* 102.4 (Oct. 2009), pp. 2069–2083.

[202] M. Sigman, G. A. Cecchi, C. D. Gilbert & M. O. Magnasco. "On a common circle: natural scenes and Gestalt rules". eng. In: *Proceedings of the National Academy of Sciences of the United States of America* 98.4 (Feb. 2001), pp. 1935–1940.

[203] A. M. Sillito, K. L. Grieve, H. E. Jones, J. Cudeiro & J. Davis. "Visual cortical mechanisms detecting focal orientation discontinuities". eng. In: *Nature* 378.6556 (Nov. 1995), pp. 492–496.

[204] E. P. Simoncelli & D. J. Heeger. "A model of neuronal responses in visual area MT". eng. In: *Vision Research* 38.5 (Mar. 1998), pp. 743–761.

[205] E. P. Simoncelli & B. A. Olshausen. “Natural image statistics and neural representation”. eng. In: *Annual Review of Neuroscience* 24 (2001), pp. 1193–1216.

[206] E.P. Simoncelli. “Statistical models for images: compression, restoration and synthesis”. In: vol. 1. IEEE Comput. Soc, 1997, pp. 673–678.

[207] E.P. Simoncelli, W.T. Freeman, E.H. Adelson & D.J. Heeger. “Shiftable multiscale transforms”. In: *IEEE Transactions on Information Theory* 38.2 (Mar. 1992), pp. 587–607.

[208] Karen Simonyan, Andrea Vedaldi & Andrew Zisserman. “Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps”. In: *arXiv:1312.6034 [cs]* (Dec. 2013).

[209] Lawrence C. Sincich & Jonathan C. Horton. “Divided by cytochrome oxidase: a map of the projections from V1 to V2 in macaques”. eng. In: *Science (New York, N.Y.)* 295.5560 (Mar. 2002), pp. 1734–1737.

[210] Lawrence C. Sincich & Jonathan C. Horton. “THE CIRCUITRY OF V1 AND V2: Integration of Color, Form, and Motion”. In: *Annual Review of Neuroscience* 28.1 (2005), pp. 303–326.

[211] A. M. Smith, B. K. Lewis, U. E. Ruttimann, F. Q. Ye, T. M. Sinnwell, Y. Yang, J. H. Duyn & J. A. Frank. “Investigation of low frequency drift in fMRI signal”. eng. In: *NeuroImage* 9.5 (May 1999), pp. 526–533.

[212] Matthew A. Smith, Najib J. Majaj & J. Anthony Movshon. “Dynamics of motion signaling by neurons in macaque area MT”. eng. In: *Nature Neuroscience* 8.2 (Feb. 2005), pp. 220–228.

[213] Alexandra Smolyanskaya, Ralf M. Haefner, Stephen G. Lomber & Richard T. Born. “A Modality-Specific Feedforward Component of Choice-Related Activity in MT”. eng. In: *Neuron* 87.1 (July 2015), pp. 208–219.

[214] Dan D. Stettler, Aniruddha Das, Jean Bennett & Charles D. Gilbert. “Lateral connectivity and contextual interactions in macaque primary visual cortex”. eng. In: *Neuron* 36.4 (Nov. 2002), pp. 739–750.

[215] Wenzhi Sun, Zhongchao Tan, Brett D. Mensh & Na Ji. “Thalamus provides layer 4 of primary visual cortex with orientation- and direction-tuned inputs”. eng. In: *Nature Neuroscience* 19.2 (Feb. 2016), pp. 308–315.

[216] Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow & Rob Fergus. “Intriguing properties of neural networks”. In: *arXiv:1312.6199 [cs]* (Dec. 2013).

[217] Seiji Tanabe & Bruce G. Cumming. “Mechanisms underlying the transformation of disparity signals from V1 to V2 in the macaque”. eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 28.44 (Oct. 2008), pp. 11304–11314.

[218] Hiroki Tanaka & Izumi Ohzawa. “Surround suppression of V1 neurons mediates orientation-based representation of high-order visual features”. eng. In: *Journal of Neurophysiology* 101.3 (Mar. 2009), pp. 1444–1462.

[219] X. Tao, B. Zhang, E. L. Smith, S. Nishimoto, I. Ohzawa & Y. M. Chino. “Local sensitivity to stimulus orientation and spatial frequency within the receptive fields of neurons in visual area 2 of macaque monkeys”. eng. In: *Journal of Neurophysiology* 107.4 (Feb. 2012), pp. 1094–1110.

[220] O. M. Thomas, B. G. Cumming & A. J. Parker. “A specialization for relative disparity in V2”. eng. In: *Nature Neuroscience* 5.5 (May 2002), pp. 472–478.

[221] M. G. Thomson, D. H. Foster & R. J. Summers. “Human sensitivity to phase perturbations in natural images: a statistical framework”. eng. In: *Perception* 29.9 (2000), pp. 1057–1069.

[222] Gasper Tkacik, Jason S. Prentice, Jonathan D. Victor & Vijay Balasubramanian. “Local statistics in natural scenes predict the saliency of synthetic textures”. eng. In: *Proceedings of the National Academy of Sciences of the United States of America* 107.42 (Oct. 2010), pp. 18149–18154.

[223] D. J. Tolhurst, Y. Tadmor & T. Chao. “Amplitude spectra of natural images”. eng. In: *Ophthalmic & Physiological Optics: The Journal of the British College of Ophthalmic Opticians (Optometrists)* 12.2 (Apr. 1992), pp. 229–232.

[224] D. J. Tolhurst & I. D. Thompson. “On the variety of spatial frequency selectivities shown by neurons in area 17 of the cat”. eng. In: *Proceedings of the Royal Society of London. Series B, Biological Sciences* 213.1191 (Oct. 1981), pp. 183–199.

[225] R. B. Tootell & S. L. Hamilton. “Functional anatomy of the second visual area (V2) in the macaque”. eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 9.8 (Aug. 1989), pp. 2620–2644.

[226] R. B. Tootell, M. S. Silverman, R. L. De Valois & G. H. Jacobs. “Functional organization of the second cortical visual area in primates”. eng. In: *Science (New York, N.Y.)* 220.4598 (May 1983), pp. 737–739.

[227] Jon Touryan, Gidon Felsen & Yang Dan. “Spatial structure of complex cell receptive fields measured with natural images”. eng. In: *Neuron* 45.5 (Mar. 2005), pp. 781–791.

[228] Doris Y. Tsao & Margaret S. Livingstone. “Mechanisms of face perception”. eng. In: *Annual Review of Neuroscience* 31 (2008), pp. 411–437.

[229] Takanori Uka & Gregory C. DeAngelis. “Linking Neural Representation to Function in Stereoscopic Depth Perception: Roles of the Middle Temporal Area in Coarse versus Fine Disparity Discrimination”. en. In: *The Journal of Neuroscience* 26.25 (June 2006), pp. 6791–6802.

[230] L. G. Ungerleider & J. V. Haxby. “’What’ and ’where’ in the human brain”. eng. In: *Current Opinion in Neurobiology* 4.2 (Apr. 1994), pp. 157–165.

[231] L. G. Ungerleider & M. Mishkin. “Two cortical visual systems”. In: *Analysis of Visual Behavior* (1982), pp. 549–586.

[232] Leslie G. Ungerleider, Thelma W. Galkin, Robert Desimone & Ricardo Gattass. “Cortical connections of area V4 in the macaque”. eng. In: *Cerebral Cortex (New York, N.Y.: 1991)* 18.3 (Mar. 2008), pp. 477–499.

[233] Laurens Van der Maaten & Geoffrey Hinton. “Visualizing data using t-SNE”. In: *Journal of Machine Learning Research* 9.2579-2605 (2008), p. 85.

[234] J. D. Victor & R. M. Shapley. “The nonlinear pathway of Y ganglion cells in the cat retina”. eng. In: *The Journal of General Physiology* 74.6 (Dec. 1979), pp. 671–689.

[235] Jonathan D. Victor & Mary M. Conte. “Local image statistics: maximum-entropy constructions and perceptual salience”. eng. In: *Journal of the Optical Society of America. A, Optics, Image Science, and Vision* 29.7 (July 2012), pp. 1313–1345.

[236] W. E. Vinje & J. L. Gallant. “Sparse coding and decorrelation in primary visual cortex during natural vision”. eng. In: *Science (New York, N.Y.)* 287.5456 (Feb. 2000), pp. 1273–1276.

[237] William E. Vinje & Jack L. Gallant. “Natural stimulation of the nonclassical receptive field increases information transmission efficiency in V1”. eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 22.7 (Apr. 2002), pp. 2904–2915.

[238] B. Vintch, E. P. Simoncelli & J. A. Movshon. “A spatial subunit model for V2 receptive fields reveals heterogeneous receptive field structure”. In: *Annual Meeting, Society for Neuroscience* (2011).

[239] Brett Vintch. “Structured hierarchical models for neurons in the early visual system”. PhD thesis. Center for Neural Science, New York University, 2013.

[240] Brett Vintch, J. Anthony Movshon & Eero P. Simoncelli. “A Convolutional Subunit Model for Neuronal Responses in Macaque V1”. eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 35.44 (Nov. 2015), pp. 14829–14841.

[241] Brett Vintch, Andrew D. Zaharia, J. Anthony Movshon & Eero P. Simoncelli. “Efficient and direct estimation of a neural subunit model for sensory coding”. ENG. In: *Advances in Neural Information Processing Systems* 25 (Dec. 2012), pp. 3113–3121.

[242] Rufin Vogels & Irving Biederman. “Effects of illumination intensity and direction on object coding in macaque inferior temporal cortex”. eng. In: *Cerebral Cortex (New York, N.Y.: 1991)* 12.7 (July 2002), pp. 756–766.

[243] Angela Voyles, Corey M Ziemba, Najib J Majaj, J Anthony Movshon & L Kiorpes. “Development of texture perception in infant monkeys: Physiology and behavior”. In: *Annual Meeting, Society for Neuroscience* (2014).

[244] G. A. Walker, I. Ohzawa & R. D. Freeman. “Asymmetric suppression outside the classical receptive field of the visual cortex”. eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 19.23 (Dec. 1999), pp. 10536–10553.

[245] Pascal Wallisch & J. Anthony Movshon. “Structure and function come unglued in the visual cortex”. eng. In: *Neuron* 60.2 (Oct. 2008), pp. 195–197.

[246] Brian A. Wandell, Serge O. Dumoulin & Alyssa A. Brewer. “Visual Field Maps in Human Cortex”. English. In: *Neuron* 56.2 (Oct. 2007), pp. 366–383.

[247] Ben S. Webb, Neel T. Dhruv, Samuel G. Solomon, Chris Tailby & Peter Lennie. “Early and late mechanisms of surround suppression in striate cortex of macaque”. eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 25.50 (Dec. 2005), pp. 11666–11675.

[248] F. A. Wichmann & N. J. Hill. “The psychometric function: I. Fitting, sampling, and goodness of fit”. eng. In: *Perception & Psychophysics* 63.8 (Nov. 2001), pp. 1293–1313.

[249] T. N. Wiesel & D. H. Hubel. “Spatial and chromatic interactions in the lateral geniculate body of the rhesus monkey”. eng. In: *Journal of Neurophysiology* 29.6 (Nov. 1966), pp. 1115–1156.

[250] Ben D. B. Willmore, Ryan J. Prenger & Jack L. Gallant. “Neural representation of natural images in visual area V2”. eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 30.6 (Feb. 2010), pp. 2102–2114.

[251] Klaus Wimmer, Albert Compte, Alex Roxin, Diogo Peixoto, Alfonso Renart & Jaime de la Rocha. “Sensory integration dynamics in a hierarchical network explains choice probabilities in cortical area MT”. en. In: *Nature Communications* 6 (Feb. 2015).

[252] Yukako Yamane, Eric T. Carlson, Katherine C. Bowman, Zhihong Wang & Charles E. Connor. “A neural code for three-dimensional object shape in macaque inferotemporal cortex”. eng. In: *Nature Neuroscience* 11.11 (Nov. 2008), pp. 1352–1360.

[253] Daniel L. K. Yamins & James J. DiCarlo. “Using goal-driven deep learning models to understand sensory cortex”. eng. In: *Nature Neuroscience* 19.3 (Feb. 2016), pp. 356–365.

[254] Daniel L. K. Yamins, Ha Hong, Charles F. Cadieu, Ethan A. Solomon, Darren Seibert & James J. DiCarlo. “Performance-optimized hierarchical models predict neural responses in higher visual cortex”. eng. In: *Proceedings of the National Academy of Sciences of the United States of America* 111.23 (June 2014), pp. 8619–8624.

[255] Jason Yosinski, Jeff Clune, Anh Nguyen, Thomas Fuchs & Hod Lipson. “Understanding Neural Networks Through Deep Visualization”. In: *arXiv:1506.06579 [cs]* (June 2015).

[256] Yunguo Yu, Anita M. Schmid & Jonathan D. Victor. “Visual processing of informative multipoint correlations arises primarily in V2”. eng. In: *eLife* 4 (2015), e06604.

[257] Andrew Zaharia, Robbe Goris, J. Movshon & Eero Simoncelli. “Compound stimuli reveal velocity separability of spatiotemporal receptive fields in macaque area MT”. eng. In: *Journal of Vision* 15.12 (2015), p. 485.

[258] Matthew D. Zeiler & Rob Fergus. “Visualizing and Understanding Convolutional Networks”. en. In: *Computer Vision – ECCV 2014*. Ed. by David Fleet, Tomas Pajdla, Bernt Schiele & Tinne Tuytelaars. Lecture Notes in Computer Science 8689. Springer International Publishing, Sept. 2014, pp. 818–833.

[259] H. Zhou, H. S. Friedman & R. von der Heydt. “Coding of border ownership in monkey visual cortex”. eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 20.17 (Sept. 2000), pp. 6594–6611.

[260] C M Ziemba, J Freeman, J A Movshon & E P Simoncelli. “Selectivity and tolerance for visual texture in macaque V2”. In: *Proceedings of the National Academy of Sciences of the United States of America* (2016).

[261] Corey M Ziemba, Jeremy Freeman, Eero P Simoncelli & J Anthony Movshon. “Linking visual perception and V2 physiology by crowdsourcing psychophysics”. In: *Society for Neuroscience Annual Meeting* (2012).

[262] Corey M Ziemba, Najib J Majaj, Romesh D Kumbhani, Christopher Shooner, Luke E Hallum, Angela Voyles, Virginia Garcia-Marin, Jenna G Kelly, J Anthony Movshon & L Kiorpes. “Amblyopia reduces neuronal and perceptual sensitivity to naturalistic image structure”. In: *Annual Meeting, Society for Neuroscience* (2014).

[263] Davide Zoccolan, Minjoon Kouh, Tomaso Poggio & James J. DiCarlo. “Trade-off between object selectivity and tolerance in monkey inferotemporal cortex”. eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 27.45 (Nov. 2007), pp. 12292–12307.